



Universidad  
Carlos III de Madrid  
[www.uc3m.es](http://www.uc3m.es)

**Máster Universitario en Bibliotecas y Servicios de Información Digital**

**2015/2016**

***Trabajo Fin de Máster***

# Journal Article Tag Suite (JATS): situación actual y análisis de implantación del estándar promovido por NISO para revistas científicas basado en XML

---

Andrea Cristiani Sienra

Tutor

José Tomás Nogales Flores

Madrid, julio de 2016

**Palabras clave:** TECNOLOGÍAS DE MARCADO; REVISTAS CIENTÍFICAS; ESTÁNDAR ABIERTO; JATS

**Resumen:** JATS es un estándar abierto ampliamente utilizado, que cuenta actualmente con una comunidad de usuarios que incluye importantes editoriales académicas, proyectos de acceso abierto a revistas científicas, bibliotecas y archivos en todo el mundo. Se describen la naturaleza, evolución y diseño del estándar JATS. Se explora el estado actual de implantación de JATS como lenguaje de marcado y su utilización como opción estandarizada para la gestión de los procesos de publicación de revistas científicas, el intercambio de información, la generación de otros productos y la preservación digital del contenido de las revistas. Se discuten las diferentes formas de implantación del estándar y de sus personalizaciones disponibles en abierto. Las conclusiones presentan los aspectos relevantes respecto a la implantación y la utilización de JATS en la actualidad.



Esta obra se encuentra sujeta a la licencia Creative Commons  
**Reconocimiento – No Comercial – Sin Obra Derivada**



## AGRADECIMIENTOS

A Tomás por su disposición y su forma de llevar a cabo la tutoría. Gracias por la orientación que abre camino al aprendizaje y a la reflexión, por las sugerencias y los comentarios siempre oportunos y muy valiosos.

A la Biblioteca de la Universidad Carlos III por el muy buen servicio de acceso a los documentos y en particular a los funcionarios del servicio de Préstamo Interbibliotecario por tramitar mis solicitudes de forma rápida y eficiente.

A los responsables de los Servicios de Publicación Digital de las bibliotecas académicas que han aceptado compartir sus experiencias y opiniones, tanto a través de la encuesta en línea como del correo electrónico. Sin su colaboración el capítulo 7 de este Trabajo Fin de Máster no hubiera sido posible.

A Gustavo Fonseca, Analista de Sistemas, Proyecto SciELO FAPESP-FapUNIFESP Brasil, por responder todas mis consultas sobre SPS.

A Claire Caulfield, *Digital/Print Operations Manager* de la *British Library* por la información brindada referida a la utilización de JATS.



## TABLA DE CONTENIDOS

AGRADECIMIENTOS.....	i
ÍNDICE DE FIGURAS .....	iv
INDICE DE GRÁFICOS .....	iv
ÍNDICE DE TABLAS .....	iv
LISTA DE ACRÓNIMOS Y SIGLAS .....	v
RESUMEN Y PALABRAS CLAVE.....	vii
ABSTRACT AND KEYWORDS .....	vii
INTRODUCCIÓN .....	1
CAPÍTULO 1. OBJETO, JUSTIFICACIÓN, METODOLOGÍA Y FUENTES .....	3
1.1. Objeto.....	3
1.2. Justificación .....	4
1.3. Metodología .....	4
1.4. Fuentes .....	6
CAPÍTULO 2. CONTEXTO .....	8
2.1. Las revistas científicas electrónicas y el medio ambiente de la publicación digital.....	8
2.2. Evolución de las formas de publicación de las revistas científicas electrónicas .....	10
2.3. Lenguajes de marcado basados en XML y su aplicación en el ámbito de las revistas científicas electrónicas .....	11
CAPÍTULO 3. ANTECEDENTES, ORIGEN, EVOLUCIÓN Y NATURALEZA DE JATS .....	15
3.1. Antecedentes: normas y estándares para el marcado de revistas .....	15
3.2. Origen y evolución de JATS .....	16
3.2.1. Orígenes. Las primeras versiones de NLM DTD .....	17
3.2.2 NLM <i>Journal Archive and Interchange Tag Suite</i> como estándar de facto .....	18
3.2.3 Estándar JATS .....	19
3.3. Naturaleza de JATS .....	21
CAPÍTULO 4. LÍNEAS DE DISEÑO Y FUNCIONALIDADES.....	24
4.1. JATS: principales líneas de diseño .....	24
4.2. La <i>Suite</i> de módulos JATS .....	28
4.3. Modelos específicos: <i>Tag Sets</i> o conjuntos de etiquetas.....	33
4.3.1. <i>Journal Archive and Interchange Tag Set</i> / Conjunto de etiquetas de archivo e intercambio .....	34
4.3.2. <i>Journal Publishing Tag Set</i> / Conjunto de etiquetas de publicación .....	35
4.3.3. <i>Article Authoring Tag Set</i> / Conjunto de etiquetas de autoría.....	37
4.4. Estructura del marcado en cada <i>Tag Set</i> .....	39



CAPÍTULO 5. PERSONALIZACIONES DE JATS.....	44
5.1. Creación de nuevos conjuntos, niveles de personalización y formas de implementación.....	45
5.2. Personalizaciones para artículos .....	47
5.2.1. TaxPub XML schema.....	47
5.2.2. Jats-supmat .....	49
5.2.3. SciELO Publishing Schema (SPS) .....	50
5.3. Personalizaciones para describir otros tipos de documentos diferentes al artículo de revista .....	51
5.3.1. Book Interchange Tag Suite (BITS) .....	51
5.3.2. ISO Standards Tag Set (ISOSTS) .....	55
CAPÍTULO 6. SITUACIÓN ACTUAL DE JATS .....	58
6.1. Formas de trabajo con XML en los procesos editoriales.....	58
6.2. JATS en editoriales y archivos con acceso abierto .....	62
6.3. JATS en editoriales académicas comerciales.....	69
6.4. JATS en bibliotecas .....	73
6.4.1. Archivo, preservación y acceso .....	73
6.4.2. Servicios editoriales.....	78
CAPÍTULO 7. ENCUESTA A BIBLIOTECAS QUE OFRECEN SERVICIOS EDITORIALES .....	81
7.1. Contexto de las bibliotecas objeto de la encuesta.....	81
7.2. Presentación y análisis de las respuestas recibidas por correo electrónico .....	82
7.3. Análisis y discusión de los resultados de encuesta .....	84
CAPÍTULO 8. CONCLUSIONES Y PROYECCIONES .....	92
8.1. Conclusiones.....	92
8.2. Proyecciones .....	99
REFERENCIAS .....	101
LISTADO DE SITIOS WEB .....	108
ANEXO 1. FORMULARIO DE LA ENCUESTA.....	111
ANEXO 2. RESUMEN DE LAS RESPUESTAS.....	113
Textos correspondientes a las preguntas con opción a respuestas abiertas.....	115



## ÍNDICE DE FIGURAS

Figura 1. Evolución de JATS .....	17
Figura 2. La referencia al archivo .dtd en un documento XML marcado con JATS .....	25
Figura 3. Modelo de contenido de la entidad Abbreviated Journal Title Elements.....	26
Figura 4. Ejemplo de entidad de parámetro para sobrescritura de modelos.....	27
Figura 5. Ejemplo de entidad de parámetro para sobrescritura de grupos de elementos.....	27
Figura 6. Ejemplo de entidad de parámetro que describe una lista de atributos .....	27
Figura 7. Comparación de las estructuras del elemento raíz en los tres Tag Sets. ....	40
Figura 8. Comparación de las estructuras del elemento <front> en los tres Tag Sets.....	41
Figura 9. Comparación de las estructuras del elemento <body> en los Tag Sets .....	41
Figura 10. Comparación de las estructuras del elemento <back> en los Tag Sets .....	42
Figura 11. Estructuras de los elementos <floats-group>, <response> y <sub-article> en los Tag Sets Archivo e Intercambio y Publicación .....	42
Figura 12. XML First o XML-in.....	59
Figura 13. XML in the middle .....	60
Figura 14. XML Last o XML-out.....	60

## ÍNDICE DE GRÁFICOS

Gráfico 1. Implementación de XML en LPS .....	85
Gráfico 2. Motivos por los cuales no implementan XML en LPS.....	86
Gráfico 3. Ventajas de implementar XML en LPS.....	88
Gráfico 4. Formatos XML utilizados por las bibliotecas .....	89
Gráfico 5. Forma de trabajo con XML en el proceso editorial - LPS.....	90

## ÍNDICE DE TABLAS

Tabla 1. Cantidad de elementos y atributos empleados en cada Tag Set, JATS v.1.0 .....	22
Tabla 2. Módulos comunes de JATS DTD Suite .....	29
Tabla 3. Módulos que definen elementos y atributos en JATS DTD Suite .....	30
Tabla 4. Módulos que definen elementos y atributos MathML 2.0 y MathML 3.0 .....	31
Tabla 5. Módulos que definen el modelo tabla inspirado en XHTML.....	31
Tabla 6. Módulos que definen entidades de caracteres y notaciones (todos opcionales) .....	32
Tabla 7. Módulos específicos de Journal Archiving and Interchange DTD (v1.0 20120330) .....	34
Tabla 8. Módulos específicos de Journal Publishing DTD .....	36
Tabla 9. Módulos específicos de Article Authoring DTD .....	38
Tabla 10. Partes estructurales de un artículo con respecto a los tres Tag Sets de JATS.....	39
Tabla 11. Estructura de los módulos para creación de nuevos Tag Sets .....	46
Tabla 12. Módulos específicos de TaxPub DTD (v0.5 Jun. 2011) .....	48
Tabla 13. Módulos específicos de JATS-supmat DTD (v0.1 20130503).....	49
Tabla 14. Módulos específicos de Book Interchange DTD (v1.0 Dec. 2013).....	52
Tabla 15. Módulos específicos de ISO Standards Tag Set DTD (v1.1 Jul. 2013) .....	56



## LISTA DE ACRÓNIMOS Y SIGLAS

<b>AAP</b>	<i>Association of American Publishers</i>
<b>ACS Pubs</b>	<i>American Chemical Society Publications Division</i>
<b>ANSI</b>	<i>American National Standards Institute [Estados Unidos]</i>
<b>APC</b>	<i>Article Publication Charge</i>
<b>ASCII</b>	<i>American Standard Code for Information Interchange</i>
<b>BBS</b>	<i>Bulletin Board System</i>
<b>BIREME</b>	<i>Biblioteca Regional de Medicina</i>
<b>BITS</b>	<i>Book Interchange Tag Set</i>
<b>CMS</b>	<i>Content Management System</i>
<b>DC</b>	<i>Dublin Core</i>
<b>DPS</b>	<i>Digital Publishing Services</i>
<b>DTD</b>	<i>Document Type Definition</i>
<b>ETDs</b>	<i>Electronic Theses and Dissertations</i>
<b>EWS</b>	<i>European Workgroup for SGML</i>
<b>FAPESP</b>	<i>Fundação de Amparo à Pesquisa do Estado de São Paulo [Brasil]</i>
<b>HoBoTS</b>	<i>Hogrefe Book Tag Set</i>
<b>ISO</b>	<i>International Standards Organization [organismo internacional de normalización]</i>
<b>ISOSTS</b>	<i>ISO Standards Tag Set</i>
<b>JATS</b>	<i>Journal Article Tag Suite</i>
<b>JATS4R</b>	<i>JATS For Reuse</i>
<b>JST</b>	<i>Japan Science and Technology Agency</i>
<b>LC</b>	<i>Library of Congress [Estados Unidos]</i>
<b>LPC</b>	<i>Library Publishing Coalition</i>
<b>LPD</b>	<i>Library Publishing Directory</i>



<b>LPS</b>	<i>Library Publishing Services</i>
<b>MAJOUR</b>	<i>Modular Application for JOURnals</i>
<b>NCBI</b>	<i>National Center for Biotechnology Information</i> [U.S. National Library of Medicine]
<b>NFAIS</b>	<i>National Federation of Advanced Information Systems</i> [Estados Unidos]
<b>NIH</b>	<i>National Institutes of Health</i> [Estados Unidos]
<b>NISO</b>	<i>National Information Standards Organization</i> [organismo nacional de normalización, Estados Unidos]
<b>NLM</b>	<i>National Library of Medicine</i> [Estados Unidos]
<b>OASIS</b>	<i>Organization for the Advancement of Structured Information Standards</i>    <i>Organization for Article Standards In Science</i>
<b>PLoS</b>	<i>Public Library of Science</i>
<b>RELAX NG</b>	<i>REgular LAnguage for XML Next Generation</i>
<b>SciELO</b>	<i>Scientific Electronic Library Online</i>
<b>SGML</b>	<i>Standard Generalized Markup Language</i>
<b>SPS</b>	<i>SciELO Publishing Schema</i>
<b>SSSH</b>	<i>Simplified SGML for Serial Headers</i>
<b>STM</b>	<i>Scientific, Technical and Medical</i>
<b>TBX</b>	<i>TermBase eXchange format</i>
<b>TEI</b>	<i>Text Encoding Initiative</i>
<b>TIC</b>	<i>Tecnologías de Información y Comunicación</i>
<b>W3C</b>	<i>World Wide Web Consortium</i>
<b>WWW</b>	<i>World Wide Web</i>
<b>WYSIWYG</b>	<i>What You See Is What You Get</i>
<b>XML</b>	<i>eXtensible Markup Language</i>
<b>XSD</b>	<i>XML Schema Definition</i>



## RESUMEN Y PALABRAS CLAVE

El estudio de la implantación de JATS en particular se justifica fundamentalmente en los beneficios presentes y futuros de la adopción generalizada de un estándar abierto para revistas científicas y obedece a la importancia creciente de contar con vocabularios y estructuras comunes basados en XML como JATS, así como también a la necesidad de conocer los cambios que actualmente están teniendo lugar en los procesos de producción, acceso y difusión al contenido de las revistas, particularmente los cambios para adaptarse al medio digital. Se describen la naturaleza, evolución y el diseño del estándar, pero también se explora el estado actual de la implantación de JATS como lenguaje de marcado, y su utilización como opción estandarizada para la gestión de los procesos de publicación de revistas científicas, el intercambio de información, la generación de otros productos y la preservación digital del contenido de las revistas. Si bien es un estándar nacional de Estados Unidos, cuenta actualmente con una comunidad de usuarios que incluye importantes editoriales académicas, proyectos de acceso abierto a revistas científicas, bibliotecas y archivos de diferentes países del mundo. Se discuten las diferentes formas de implantación del estándar y de sus personalizaciones disponibles en acceso abierto, por parte de diversos tipos de emprendimientos, desde editoriales académicas comerciales a proyectos que apuestan al acceso abierto, incluidas en esta última categoría la mayoría de las bibliotecas académicas que ofrecen servicios editoriales a sus comunidades. Las conclusiones presentan aspectos relevantes respecto a la implantación y la utilización de JATS en la actualidad.

TECNOLOGÍAS DE MARCADO; REVISTAS CIENTÍFICAS; ESTÁNDAR ABIERTO; JATS

## ABSTRACT AND KEYWORDS

The study of the implementation of JATS in particular is mainly justified on the present and future benefits that the widespread adoption of an open standard has for scientific journals. It reflects the growing importance that common vocabularies and XML-based structures such as JATS have, as well as the need to understand the changes that are currently taking place in production processes, access and dissemination of journal contents, particularly the changes to adapt to the digital environment. Nature, evolution and design of the standard are described, as well as the current state of implementation of JATS as markup language, and its use as a standardized option for managing scientific journal publishing processes, generating other products, information exchange and digital preservation of journal content. Although JATS is a national US standard, it currently has a community of users that includes major commercial academic publishers, open access journals, libraries and archives in several countries around the world. The diverse ways of implementing the standard and its publicly available customizations by different kinds of enterprises, are discussed: from commercial academic publishers to those that are committed to open access. Most academic libraries offering publishing services to their communities are included in the latter category. The conclusions remark important aspects regarding nowadays implementation and usage of JATS.

MARKUP TECHNOLOGIES; SCIENTIFIC JOURNALS; OPEN STANDARD; JATS



## INTRODUCCIÓN

En 2015 se han cumplido 350 años desde la aparición de las primeras revistas científicas. Fue en 1665 que el *Journal des Sçavans y Philosophical Transactions* de la *Royal Society* de Londres se publicaron por primera vez, en Francia e Inglaterra, respectivamente. Las revistas han continuado formando parte fundamental del proceso de comunicación académica durante todo este tiempo, y constituyen una parte integral de la propia investigación científica, pues no solo difunden la información, sino que mantienen la calidad de los contenidos a través de mecanismos basados en la revisión por pares, siendo además fuente y registro de información para otros científicos, actuales y futuros.

En la era de las tecnologías de la información, las revistas científicas no han permanecido ajenas a la creciente ola de avance tecnológico. Los cambios tecnológicos han modificado las formas de comunicación del conocimiento y cómo nos vinculamos con él. Las revistas han cambiado las formas en que muestran sus contenidos, y la web se vuelve cada vez más importante como forma de difusión de los mismos.

Se pueden identificar tres formatos que coexisten actualmente en la distribución y el acceso a revistas científicas: formato impreso, impreso y en línea y sólo en línea (si bien se ha dado una evolución en el tiempo del primero hacia el segundo y de éste hacia el tercero). El formato solo en línea satisface muy bien la demanda creciente por una rápida publicación de los resultados de investigación. De esta forma, la publicación en formato digital gana terreno frente a otras formas de publicación.

La motivación por estudiar este tema tiene como punto de partida un interés personal y profesional por conocer los cambios que se están dando actualmente en los procesos de producción, acceso y difusión de las revistas, particularmente los cambios para adaptarse al medio digital. Esos cambios han implicado entre otras cosas, la adopción de esquemas de marcado basados en uno de los metalenguajes estándar para la *World Wide Web*: el XML. JATS es uno de los esquemas actuales basados en XML para marcado estructural y semántico de los contenidos de revistas. No es el único, pero se constituye en una muy buena opción al ser un estándar abierto, no propietario, desarrollado y mantenido por un organismo nacional de normalización, ya que es estándar nacional ANSI/NISO en Estados Unidos desde el año 2012. Cuando planteamos inicialmente el tema del trabajo, la principal motivación por el estudio de la implantación de JATS en particular se focalizó fundamentalmente en los beneficios presentes y futuros que se pueden avizorar en la adopción generalizada de un estándar para revistas científicas.

Coexisten actualmente variados modelos de negocio para la producción y distribución de revistas científicas, y diferentes flujos o procesos de trabajo editorial. Si bien en muchos casos se utilizan lenguajes y vocabularios propios, la tendencia es ir hacia vocabularios comunes que faciliten y mejoren los procesos internos, así como también los servicios ofrecidos a los usuarios. La importancia creciente de contar con vocabularios y estructuras comunes basados en XML como JATS obedece a razones que van desde facilitar el tan necesario intercambio de contenido digital, hasta la propia preservación digital, que permitirá poder acceder y hacer uso de esos contenidos en el futuro.

Este es un trabajo de carácter descriptivo de la naturaleza y el diseño del estándar, pero también exploratorio, pues indaga sobre la situación actual de implantación de JATS. Para llegar a comprender cabalmente la situación actual, se parte de un estudio de la evolución que ha tenido JATS, incluso desde antes de convertirse en estándar. Además, para comprender el estado actual de



implantación es necesario conocer los detalles sobre su diseño original y las diversas personalizaciones que se le han realizado para adaptarlo a diferentes requerimientos editoriales. Por ello dedicamos un capítulo a la descripción del diseño original y otro a las personalizaciones de JATS que están disponibles para su descarga de forma pública.

El trabajo está dividido en 7 capítulos, cuya secuencia, sin perder de vista los objetivos del trabajo y teniendo en cuenta un desarrollo ordenado, busca comunicar al lector la información de manera clara, introduciendo primero el contexto, luego los antecedentes, la evolución y la naturaleza del estándar, para luego presentar algunas personalizaciones de JATS y finalmente centrarse en la situación actual de implantación de JATS.

En el capítulo 1 se explica la propuesta de investigación, incluyendo la demarcación del objeto de estudio, la justificación de la propuesta y los objetivos que se pretenden alcanzar, así como también la metodología y las fuentes utilizadas para alcanzarlos. El capítulo 2 introduce el contexto de la publicación de revistas científicas y de las tecnologías de marcado basadas en XML. El capítulo 3, describe la evolución histórica de JATS, la cual es muy importante si pretendemos entender mejor su presente. El capítulo 4 trata sobre las principales líneas de diseño del estándar, e incluye una descripción de la *Tag Suite* y los tres conjuntos de etiquetas creados a partir de ella. El capítulo 5 está dedicado a presentar algunas de las personalizaciones de JATS (nuevos conjuntos de etiquetas basados en el estándar) utilizadas actualmente que están disponibles en acceso abierto. El capítulo 6 se enfoca en la situación actual de implantación del estándar, valiéndose para ello de una selección de casos concretos de utilización de JATS o alguna de sus personalizaciones. Dentro del panorama actual de implantación, se incluye la situación de las bibliotecas respecto a si utilizan o están pensando utilizar JATS en sus productos y servicios. En último lugar, en el capítulo 7 se presentan las conclusiones finales y las posibles proyecciones de este trabajo.

## CAPÍTULO 1. OBJETO, JUSTIFICACIÓN, METODOLOGÍA Y FUENTES

En este capítulo se presentan el objeto de estudio del Trabajo Fin de Máster, los objetivos propuestos y la justificación del tema escogido, conjuntamente con algunas preguntas que se pretenden responder concluido el estudio. Luego se explica la metodología adoptada para alcanzar los objetivos propuestos y finalmente se mencionan las principales fuentes consultadas para llevar a cabo el trabajo.

### 1.1. Objeto

El objeto de estudio es la implantación del estándar NISO JATS para la producción y gestión de contenidos de revistas científicas, sean éstos de carácter digital disponibles en línea a través de Internet o impresos.

El estándar JATS es un esquema basado en XML, cuya última versión es la 1.1 de noviembre de 2015. Si bien el comité que mantiene el estándar y los grupos de trabajo están integrados por expertos de distintas nacionalidades y provienen de diferentes instituciones y organizaciones con o sin fines de lucro, no es un estándar internacional, es un estándar nacional promovido por *National Information Standards Organization* (NISO), un organismo de estandarización en materia de información de los Estados Unidos. Cabe entonces preguntarse qué tanto ha logrado trascender fuera de ese territorio, quiénes lo implementan en la actualidad, y sobre todo qué uso le dan quienes lo implementan y cómo lo hacen ¿"tal cual es" o con modificaciones? Resulta interesante conocer también qué tan extendido está en el ámbito de las bibliotecas: ¿las bibliotecas lo implementan?, ¿les interesa como opción para desarrollar productos y servicios?, ¿por qué motivos lo utilizan / no lo utilizan?

Se quiere investigar el estado de implantación actual de JATS, quiénes lo utilizan, cómo lo implementan y para qué fines. Se intenta descubrir si existe una relación entre el modelo de negocio editorial y los usos específicos que se pueden apreciar en las formas de implantación existentes ("tal cual es", parcial o personalizada).

#### **Objetivo general:**

Realizar un análisis de la evolución y el estado actual de la implantación de JATS como lenguaje de marcado, y su utilización como opción estandarizada para la gestión de los procesos de publicación de revistas científicas, el intercambio de información, la generación de otros productos y la preservación digital del contenido de las mismas.

#### **Objetivos específicos:**

1) Estudiar las líneas principales que sustentan el diseño de JATS, desde sus antecedentes hasta la última versión oficial: v 1.1 (ANSI/NISO Z39.96-2015).

2) Investigar los usos específicos actuales que se sustentan en la aplicación de JATS y su utilización como opción estandarizada para la gestión de los procesos de publicación, la elaboración de productos, el intercambio de información y la preservación digital del contenido de revistas científicas, y otros tipos de documentos diferentes al contenido de revista que utilicen JATS como base en sus esquemas.

3) Aportar un panorama actual de la implantación de JATS en bibliotecas, principalmente enfocado a explorar si implementan o están planeando implementar JATS para la prestación de servicios editoriales desde la biblioteca.

## 1.2. Justificación

La elección de investigar y profundizar en el tema estándares de marcado tiene que ver en parte con los intereses y el gusto personal, además de la intención de formar un perfil profesional en esta área de especialización. El interés que despertaron y los conocimientos que aportaron a mi formación varias asignaturas de este máster, especialmente Tecnologías de Marcado para Textos Digitales, pero también Vocabularios y Esquemas Semánticos para la Web y Digitalización y Preservación, también tuvieron mucha influencia en la elección de este tema.

El aspecto práctico de la realización de este estudio se fundamenta en los beneficios de aportar información actual sobre la implantación de un estándar que está siendo utilizado con éxito en muchos proyectos editoriales e institucionales de la más diversa envergadura, en portales de revistas que proveen servicios de información basados en contenidos científicos y académicos en distintos países del mundo. Utilizar un lenguaje basado en XML o no, es una decisión que se toma al momento de planificar el producto o servicio, y que se fundamenta en la necesidad de optar por tecnologías, mecanismos y procedimientos para crear ese producto o servicio rápida y efectivamente, hacerlo disponible en distintas plataformas y dispositivos y además preservar el contenido completo de los documentos para el futuro. Un trabajo que describe, analiza y actualiza la información sobre el estándar JATS, cuya existencia como tal tiene relativamente pocos años, tiene un valor potencial importante en contextos donde se deba tomar una decisión de este tipo.

Los destinatarios más directos de los resultados de este trabajo se encuentran principalmente en el ámbito de los servicios de publicación digital para revistas académicas, tanto aquellos que brinden esos servicios actualmente como aquellos que estén pensando brindarlos, o bien dar asesoramiento en esa área. En ese grupo de interés están incluidas las bibliotecas, especialmente las bibliotecas de universidades e institutos de investigación, los editores de revistas académicas, los profesionales bibliotecarios y los estudiantes de biblioteconomía, los nuevos proyectos y emprendimientos editoriales (o los antiguos que estén considerando incluir tecnologías basadas en XML), los interesados en publicación de revistas científicas en general y en estándares de marcado para revistas científicas en particular.

Un trabajo de estas características resulta necesario para reunir y analizar la información disponible sobre el estado actual de la implantación de este estándar y además cumple con el cometido de hacerla accesible a todos aquellos interesados en la temática que siendo hispanohablantes no acceden a la información más reciente relacionada con el estándar JATS, pues mucha de ella se expone en conferencias internacionales en inglés y no está disponible en idioma español.

## 1.3. Metodología

La metodología utilizada es diferente según el objetivo específico, siendo el abordaje descriptivo para el primer objetivo y utilizando luego un enfoque exploratorio para el segundo y el tercero.

En el caso del primer objetivo, para presentar las líneas de diseño y su evolución, se realiza un estudio basado en la recopilación de información mediante revisión bibliográfica de recursos que permiten conocer la evolución del estándar, sus antecedentes y cómo fue pensado su diseño. La revisión detallada de la documentación del propio estándar es un método que aporta muchos elementos al conocimiento de las líneas de diseño, así como también las notas incluidas en el código fuente de las *Document Type Definition* (DTD) de los tres conjuntos de etiquetas incluidos en el estándar. Para ello se descargaron los esquemas de los tres conjuntos de etiquetas, y se accedió a la DTD de base y a cada uno de los módulos de cada conjunto. Las anotaciones y los comentarios en el código resultan muy útiles para entender el diseño y el funcionamiento de JATS.

Del estudio de la documentación de JATS y de las notas en las DTD de base y sus módulos resultaron varias tablas que contienen una síntesis de la información sobre los componentes de JATS y sus respectivas funciones, las cuales presentamos en los capítulos 4 y 5. Con este método se estudiaron las estructuras de los módulos, tanto en el caso del diseño original del estándar como para algunas de sus extensiones. Las tablas resultaron un recurso útil para presentar las estructuras de cada conjunto de etiquetas JATS de manera clara y concisa en el capítulo 4, incorporando a ellas además otros datos como los nombres de los módulos y de sus respectivos ficheros, y la función específica de cada uno. Las tablas del capítulo 5 sirven para visualizar rápidamente la ampliación de la estructura modular en las DTD extendidas, pues además de presentar la estructura básica para la creación de nuevos conjuntos de etiquetas a partir de los módulos del estándar, se ha agregado al final información sobre los módulos que se definen exclusivamente para los nuevos modelos de contenido.

Para la consecución del segundo objetivo específico, el cual explorara el panorama actual, se ha realizado un estudio exploratorio basado en escoger, presentar y analizar casos concretos de bibliotecas, editoriales, portales de revistas científicas y otros productores de servicios y/o de recursos de información que utilicen JATS o alguna de sus personalizaciones como tecnología de marcado.

Explorar la situación actual implica buscar e indagar sobre casos concretos de uso, para lo cual se han investigado distintos tipos de emprendimientos llevados a cabo por diferentes instituciones y organizaciones, desde editoriales con diferentes modelos de negocio hasta bibliotecas que implementen JATS. Para ello se toman algunos casos específicos de servicios o proyectos existentes, de forma de representar con esa selección la variedad de opciones de implantación de JATS en la actualidad. Este abordaje exploratorio conlleva una metodología basada en la recopilación de información sobre ejemplos actuales de utilización del estándar JATS, mediante la consulta de sitios web, publicaciones académicas, conferencias de expertos, comunicaciones personales, etc.

Para el tercer objetivo de explorar la situación de implantación de JATS en bibliotecas, y especialmente en aquellas que ofrecen servicios de publicación, además de la recopilación de información basada en la bibliografía, se utilizó la técnica de la encuesta y una metodología cuantitativa de análisis de los datos. Los aspectos específicos del diseño de la herramienta de recolección de datos (formulario de la encuesta) y los criterios para distribuir la encuesta se explican en el capítulo 7.

La encuesta que se diseñó para este estudio tiene como interés principal conocer la situación actual de implantación de JATS en bibliotecas académicas. Se distribuyó mediante un formulario en línea destinado a los responsables de servicios de publicación digital. La herramienta busca recabar con las respuestas un testimonio actual basado en las experiencias y opiniones de los encuestados.

## 1.4. Fuentes

Se ha recurrido a varios tipos de fuentes para cubrir los distintos objetivos que nos planteamos. Se han utilizado tanto fuentes que proporcionan una visión general del estado actual de implantación, como documentos que proporcionan información y datos concretos sobre la evolución histórica, tanto del tema estudiado como de su contexto: el medio ambiente de la publicación digital de revistas científicas.

Las fuentes específicas de información sobre la naturaleza y el diseño de JATS han sido varias. La principal fuente de información que se ha consultado es la documentación oficial del estándar en el sitio web de NISO. También se recurrió a documentos disponibles en el sitio web de *National Library of Medicine* (NLM) dedicado a JATS y sus antecesores, además de toda la información adicional proporcionada por la Wiki oficial de JATS. Todos esos sitios abordan aspectos conceptuales relacionados con definiciones, diseño y evolución del estándar, así como también tratan aspectos prácticos de su implantación, brindando acceso a herramientas y aplicaciones para producir y validar documentos XML de acuerdo al estándar y enlaces directos para la descarga de los conjuntos de etiquetas.

Otras fuentes como las conferencias, seminarios y simposios de expertos han resultado de particular importancia para comprender la situación actual de implantación de JATS, así como también para obtener datos recientes sobre lenguajes de marcado basados en XML relacionados con el ambiente de la publicación digital. Los eventos que se han tomado como fuente son, en primer lugar *Journal Article Tag Suite Conference (JATS-Con) Proceedings* por la especificidad de la temática de esta conferencia en relación con el tema desarrollado en este trabajo, y en segundo lugar otros eventos relacionados con la industria editorial, o con el desarrollo de aplicaciones XML como por ejemplo *XML Prague, International Symposium on Versioning XML Vocabularies and Systems*, *International Symposium on XML for the Long Haul: Issues in the Long-term Preservation of XML*, *E-production Seminar (International Association of STM Publishers)*, *International Symposium on Quality Assurance and Quality Control in XML*.

Han sido muy importantes al momento de elaborar un marco de referencia sobre el tema los artículos en revistas especializadas como *Journal of Electronic Publishing*, *Information Standards Quarterly*, *Balysage Series on Markup Technologies*, *OCLC Systems and Services: International digital library perspectives*, *Library Hi Tech*, *Library Trends*, *Scholarly and Research Communication*, *Journal of the Association for Information Science y Technology*, *Portal: Libraries and the Academy*.

Diversos sitios web pertenecientes a diferentes portales de editoriales científicas de distintos países han sido consultados como fuente de información acerca de la implantación de JATS en diversos productos: producción editorial de revistas y libros, generación de indicadores bibliométricos y cienciométricos, agregación de contenidos, etc. El listado de los sitios web consultados está disponible al final del trabajo, a continuación de las referencias bibliográficas.

En cuanto a las fuentes para los aspectos metodológicos, uno de los métodos utilizados para alcanzar el objetivo 3 es cuantitativo y se vale de la técnica de la encuesta. Se utilizó el libro de Padua y otros (1996) titulado “Técnicas de investigación aplicadas a las ciencias sociales” como guía en el diseño de la encuesta. Es un manual de técnicas de investigación que centra su interés en los métodos de *survey*, o investigaciones por cuestionarios y su modo de construcción y aplicación.

Han surgido durante el transcurso de la investigación algunas cuestiones que no se pudieron resolver con la información disponible a través de las fuentes documentales, por lo cual además se ha

recurrido a otras fuentes: listas de discusión y contactos vía correo-e. Se han realizado consultas en una lista de discusión especializada (Lista de discusión SciELO XML de *Google Groups*, administrada por el Proyecto SciELO), con el objeto de despejar algunas dudas concretas respecto a una de las personalizaciones de JATS, además de contactos personales a través de correo-e para investigar el uso de JATS en la *British Library*.

También se han consultado fuentes de referencia, como *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists* y *Library Publishing Directory*. *Library Publishing Directory* es una publicación impresa y un recurso en línea, editado por *Library Publishing Coalition* (LPC), que presenta información y datos actuales sobre tendencias en servicios editoriales ofrecidos por bibliotecas. Se consultaron las ediciones 2014, 2015 y 2016. La lista de contactos que figura en la edición 2016 del directorio resultó muy adecuada al momento de distribuir la encuesta, por lo específico del alcance del directorio, y por su actualidad. Esto último es muy significativo al momento de valorar la vigencia de la información allí contenida.



## CAPÍTULO 2. CONTEXTO

Este estudio se enmarca en la realidad actual de la publicación de contribuciones científicas y académicas, esencialmente en el caso de las revistas científicas. Por ese motivo, en este capítulo realizamos una breve introducción sobre qué son las revistas científicas y las revistas científicas electrónicas en particular, para luego contextualizarlas en el medio ambiente de la publicación digital. Las formas de distribución, producción y preservación de las revistas, cada vez más orientadas a la web, conllevan nuevas opciones y elecciones de tecnologías orientadas a ese medio, y es allí donde los estándares basados en lenguajes de la web logran su mejor desempeño.

Para finalizar este capítulo sobre el contexto de implantación de JATS, presentamos una breve síntesis de la utilización de lenguajes basados en SGML primero, y posteriormente en XML en el ámbito de las revistas científicas electrónicas. Al ser JATS un estándar basado en XML, introducimos en este capítulo sobre contexto las características principales de este metalenguaje, haciendo referencia a las ventajas, desventajas y opciones de utilización como base para el desarrollo de lenguajes de marcado para documentos en general y de contribuciones a revistas científicas en particular.

### 2.1. Las revistas científicas electrónicas y el medio ambiente de la publicación digital

La publicación de artículos de revista en formato electrónico forma parte de un contexto más amplio que es el de la publicación digital, publicación electrónica o *e-publishing*.

F.W. Lancaster (1995) reflexionaba sobre la definición de publicación electrónica y sostenía que si el término publicar implica producción y distribución, entonces obviamente el término publicación electrónica (*electronic publishing*) refiere a la generación de publicaciones en forma electrónica, o al menos con la ayuda de la electrónica. El término "revista electrónica" le resultaba casi tan ambiguo como el término "publicación electrónica", pudiendo abarcar cualquier revista existente en un formato electrónico.

Actualmente, cuando nos referimos a las **revistas electrónicas** (*e-journals*), estamos aludiendo a revistas especializadas que se publican en Internet. Este tipo de publicaciones cuentan con determinadas características que las hacen pertenecer a la categoría publicaciones seriadas (periodicidad, se publican en fascículos numerados), y pueden ser clasificadas de acuerdo a sus contenidos y al público al que van dirigidas en revistas científicas (*scientific journals*), boletines (*newsletters*), o "de divulgación" (*magazines*). Este trabajo tiene por su temática un contexto particular al que nos estaremos refiriendo repetidamente y es el de las **revistas científicas electrónicas**, y las contribuciones que en ellas se publican, siendo el artículo científico la más usual de las contribuciones. Los **artículos científicos** se caracterizan por su contenido especializado y su propósito de proveer un canal de comunicación para la investigación científica y académica.

Owens (2010) se refiere al medio ambiente de la publicación digital como dinámico, cambiante y en continua evolución. Es sin dudas un ambiente en el cual las manifestaciones y las expresiones de los contenidos informativos son muy variadas. Además de muy dinámico, está cada vez menos centrado en el formato del documento y más orientado a la información contenida en el documento. Cambian los formatos, los lenguajes, las formas de producción, los estándares, pues las necesidades y el contexto cambian y se complejizan.

Beck (2010) menciona que, debido a los rápidos cambios en las formas de publicación de los artículos desde la invención de Internet, los modelos de publicación de revistas han comenzado a cambiar. Un negocio que tenía un modelo bastante estable de artículos en números que se imprimían



y enviaban por correo postal, se ha ido desplazando hacia la publicación de forma continua, con artículos disponibles en línea tan pronto como estén listos, los cuales posteriormente son recogidos en una versión impresa, o bien nunca llegan a ser impresos. Las primeras versiones de PubMed Central fueron creadas teniendo en mente el modelo de publicación basado en números de revista, pero ese modelo ha cambiado y el modelo actual debe ser más flexible para manejar esos cambios.

Es en este ambiente de cambios dinámicos, de surgimiento y transformación de las formas de producción, publicación y distribución de las publicaciones en el que surge el estándar JATS para artículos científicos, cuyo objeto es describir un conjunto de elementos y atributos que son utilizados para construir modelos en XML para los artículos de revista. Los artículos de revistas científicas son el ámbito de aplicación del estándar JATS en la actualidad, el cual también abarca otros formatos de contribuciones a revistas, no solamente el artículo de investigación.

Tomando la analogía que propone Owens (2010) con el concepto de medio ambiente y llevándolo al ámbito de la publicación de artículos de revista, podemos identificar diferentes actores, tanto personas en su rol como individuos, como integrando comunidades de usuarios, instituciones y organizaciones. En este medio ambiente interactúan entre sí editores comerciales, portales colaborativos que publican en acceso abierto revistas científicas en diferentes áreas del conocimiento, ya sean éstos nacionales o regionales (*National Library of Medicine* (NLM), SciELO), autores, bibliotecas, profesionales de la información, ingenieros de software, editoriales, servicios de indización y resúmenes, comunidades de usuarios, y básicamente todos los que tienen algún de vinculación con la creación, edición, publicación y uso de las revistas científicas.

El editor de una revista científica es por lo general un experto en su campo, frecuentemente (pero no siempre) un académico universitario. Entre sus principales funciones está recibir los artículos de los autores para juzgar su pertinencia para la revista, y remitirlos a otros expertos para la revisión por pares. La encargada de la publicación, distribución, mercadeo y comercialización es la editorial (*Publisher*). Su rol es el proveer servicios de publicación, se encarga de editar y marcar los artículos. Si la revista tiene versión papel se encarga también de imprimir y encuadernar los ejemplares.

Entre las organizaciones que publican una o dos revistas, como los departamentos de investigación académica, las agencias gubernamentales, o las sociedades científicas, sucede a menudo que muchas de ellas ni siquiera se consideran a sí mismas como editoriales, si bien editan y publican ellos mismos sus propias revistas. En el otro extremo, están las grandes editoriales internacionales, como Pearson, Elsevier, Sage, las editoriales académicas Harvard, Toronto, Taylor y Francis, o los organismos internacionales. Hay diferencias sustanciales entre pequeñas y medianas editoriales frente a las más grandes en muchos aspectos de la producción editorial y de las tecnologías que emplean en el proceso. Esto se tratará en el capítulo 6.

Además, estos últimos años han surgido nuevos roles para algunos actores. En el caso de la biblioteca universitaria, un rol emergente consiste en brindar servicios de publicación en formato digital accesible en Internet a las facultades y departamentos de sus respectivas universidades. Esta actividad se ha denominado *Library-based publishing*. Ha surgido en el ámbito de las bibliotecas académicas estadounidenses y canadienses, como lo señalan Huwe (2014) y Lorimer (2013) pero se ha extendido rápidamente a bibliotecas de otros países como Australia, Alemania, Suecia, Escocia, Inglaterra e Irlanda como menciona Lippincott (2015). Son actualmente 115 las bibliotecas universitarias a nivel global las que integran la *Library Publishing Coalition* (LPC) y brindan estos servicios, según el directorio *Library Publishing Directory 2016*.

## 2.2. Evolución de las formas de publicación de las revistas científicas electrónicas

El formato electrónico cambió la forma de producir y distribuir revistas científicas. Han pasado ya más de 20 años desde que aparecieron los primeros artículos científicos disponibles en línea en la web. En estos años han cambiado mucho las tecnologías de publicación y las formas de tratar la información en ellos contenida.

En la primera edición de julio de 1991 del *Directory of Electronic Journals, Newsletters, and Academic Discussion Lists, Volume 1* de la *Association of Research Libraries* de Washington y refiriéndose a las **revistas electrónicas** se menciona:

*It is a visible manifestation of the state-of-the-art for the continuing networked electronic writings and correspondence that many academics and librarians already describe as a new type of journal or serial: rapid, interactive, collaborative scholarly communication. The directory will demonstrate where we are at the beginning of the last decade of the 20th century.*

En esta primera edición del directorio antes mencionado, aparecen 36 referencias dentro de la categoría *Journals*, entre las cuales destaca la revista *Flora Online* editada por la Universidad de Harvard. Fue publicada por primera vez en enero de 1987, con la característica de ser *peer review*, accesible vía FTP, *Bulletin Board System* (BBS) y también disponible en diskette, en formato ASCII. Posteriormente también se hizo accesible vía Gopher, en la década de los 90. Según Lach Arlinghaus y Zander (2008), es la primera revista electrónica en recibir un ISSN de la Biblioteca del Congreso: ISSN 0892-9106.

En los 90, convivían diferentes modalidades de acceso a las revistas electrónicas: vía lista de distribución por correo electrónico, Gopher, FTP, WWW (Mosaic), y los principales formatos estándar de publicación eran ASCII, PostScript, HTML y TeX.

En los comienzos de la WWW, para muchas revistas la versión web era una copia escaneada de lo que se publicaba en papel. Lach Arlinghaus y Zander (2008) señalan que ya en 1993 se distinguía entre “*Mere computerized versions of hardcopy titles*” y “*‘genuine’ electronic journal*”. Estos autores utilizan el término “*genuine*” para referirse a las revistas electrónicas que necesitan la capacidad de las redes para utilizar todas las posibilidades que brinda el hipertexto, y que no se limitan solamente a brindar el texto completo. Cuando el texto completo de las revistas impresas se hizo accesible en línea, el texto estaba en formato ASCII permitiendo todo tipo de búsquedas. Sin embargo, este tipo de revistas eran originalmente revistas impresas a las que luego se accedía electrónicamente.

Lancaster (1995) al hablar de las revistas en formato electrónico, también considera importante hacer una distinción entre las revistas electrónicas en línea, y las revistas impresas en papel a las cuales es posible acceder en formato electrónico posteriormente, ya sea en CD-ROM o en línea. Las revistas mencionadas en primer lugar han sido diseñadas desde un principio como revistas en formato electrónico y tienen prestaciones que no están presentes en las manifestaciones electrónicas de las revistas impresas. Por ejemplo, el texto de las revistas electrónicas en línea puede ser codificado con etiquetas SGML para mejorar su funcionalidad con la aplicación de características tales como hipertexto y la integración de texto con gráficos.

Muchas editoriales aún hoy continúan publicando versiones en PDF que reproducen la apariencia exacta de la versión papel, si bien integran otras funcionalidades al PDF, o brindan acceso también a versiones en otros formatos electrónicos. No abundaremos en los formatos utilizados para la representación del contenido ya que en este trabajo interesa la forma de estructurar lógica y semánticamente un artículo, no los formatos de sus representaciones.

F.W. Lancaster (1995) realiza una síntesis de los cambios graduales en la evolución de la publicación electrónica. Divide su evolución en cuatro etapas: la primera etapa la sitúa a principios de la década del 60, con el uso de ordenadores para generar publicaciones que luego se imprimían en papel (ej.: la producción del *Index Medicus* en la *National Library of Medicine*). La segunda etapa consistió en la distribución de texto en forma electrónica, donde la versión electrónica es el equivalente exacto de la versión papel y podía o no ser usada para generar la versión impresa en papel. Para las fuentes secundarias de información (servicios de indexación y resúmenes) la distribución electrónica comenzó a principios de los 60, pero para las fuentes primarias como las revistas, comenzó algo después. La tercera etapa es la distribución en forma electrónica solamente, pero siendo esta versión poco más que la versión papel presentada de forma electrónica. Sin embargo, tiene como valor añadido varias características, que incluyen la búsqueda, la manipulación de datos y las alertas (a través de un perfil de usuario). La cuarta etapa es la generación de publicaciones que explotan las verdaderas capacidades del medio electrónico (hipertexto e hipermedia, sonido, imagen, video). Esta etapa Lancaster la subdivide en dos fases: a) la presentación de forma innovadora de textos y gráficos ya existentes y b) la producción de publicaciones diseñadas *ab initio* para explotar a pleno las capacidades del medio electrónico. Es esta necesidad de explotar todas las capacidades del medio electrónico lo que determina en parte el surgimiento de estándares de publicación para revistas electrónicas.

Las editoriales han ido desarrollando a lo largo del tiempo diferentes formas de estructurar el contenido lógico y semántico de los artículos de sus revistas electrónicas. Esta diversidad está basada en la elaboración de lenguajes de marcado propios, con diferentes estructuras y utilizando diferentes vocabularios. Se los denomina lenguajes de marcado porque insertan marcas en un documento a base de etiquetas asignadas a cada parte estructural. Estas etiquetas describen las partes lógicas (estructura lógica y semántica) de un tipo de documento o datos particular.

Si bien las aplicaciones y las plataformas de muchas de las editoriales comerciales son propietarias, éstas también han llegado a la conclusión de necesitar un formato común para la estructura lógica del contenido. En el caso de las revistas, la NLM y sus colaboradores son quienes más éxito han tenido desarrollado e implantando un lenguaje de marcado para los artículos que se ocupa del marcado lógico, a nivel estructural y semántico, codificado en XML. Este lenguaje separa contenido de presentación facilitando con ello el intercambio, la gestión, la recuperación, el almacenamiento, y la preservación de los contenidos. Se trata del estándar NISO JATS (*Journal Article Tag Suite*). Por su característica de estándar abierto, tanto editores de revistas en acceso abierto como editores comerciales lo están utilizando actualmente para marcar los artículos que producen.

### 2.3. Lenguajes de marcado basados en XML y su aplicación en el ámbito de las revistas científicas electrónicas

XML se convirtió en un estándar del *World Wide Web Consortium* (W3C) en 1998, pero la mayoría de los editores académicos comenzaron a archivar sus contenidos de forma estructurada antes de la creación de XML. En el año 2002, la mayor parte de los editores comerciales aún utilizaban DTD en SGML (Rosenblum y Golfman, 2002).

XML (por su acrónimo en inglés *eXtensible Markup Language*), Lenguaje de marcado extensible, es una versión abreviada de SGML, la cual omite las partes complejas y menos utilizadas en SGML.

*The Extensible Markup Language (XML) is a subset of SGML [...]. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with*

*HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML (Bray y otros, 2008).*

Una de las principales características de los documentos XML es que su estructura está compuesta de etiquetas, las cuales son indicadas en el marcado y es lo que posibilita que las aplicaciones de software identifiquen, recuperen y procesen el contenido de esas etiquetas.

*XML documents are structured documents where a document instance is described as a hierarchic structure of named parts. The parts are explicitly indicated by systematic markup that enables applications to identify, retrieve, and process those parts (Bray y otros, 2008).*

XML establece estructuras de tipos de documentos (o conjuntos de datos) por medio de algún lenguaje de esquemas (en sentido amplio), dos de los cuales son el de las DTD, que usan una sintaxis diferente del marcado XML, y el de las XSD (*XML Scheme Definition*) del W3C, que sí utiliza la sintaxis de marcado XML. De esta manera se crean lenguajes, vocabularios o aplicaciones que definen no sólo una lista de "elementos válidos" sino que especifican mucho más: los atributos de éstos y sus tipos, y sobre todo, fijan las construcciones estructurales que están permitidas en base a esos elementos. Las DTD tienen la particularidad que pueden declararse de forma interna o externa al documento XML. Las DTD externas son también documentos XML, con extensión .dtd.

El interés en utilizar XML como base para la creación de lenguajes de marcado a principios de la década pasada, estuvo guiado principalmente por la preocupación de las bibliotecas en preservar y recuperar a largo plazo la información contenida en las revistas electrónicas, y por el beneficio que suponía para las editoriales de revistas poder contar con esquemas de contenido que les permitieran intercambiar con facilidad, facilitar la búsqueda y recuperación, y a la vez reducir costos y tiempos de publicación acortando los procesos de producción.

Entre las tecnologías disponibles para el tratamiento informático de artículos científicos y otros objetos digitales relacionados, los desarrolladores de JATS se decantaron por el tratamiento con un lenguaje de marcado. Para decidir el metalenguaje, se basaron XML, pues a partir de 2001 se hace evidente que XML era el futuro dentro de los metalenguajes existentes en ese momento.

XML es el metalenguaje que utiliza JATS actualmente para codificar su lenguaje de marcado. Dicho de otra manera, es XML quien proporciona la sintaxis para marcar lógicamente la estructura y la semántica de los artículos que utiliza este estándar. En el texto de un artículo pueden marcarse por ejemplo el título, los autores, resumen, palabras claves, párrafos, tablas, imágenes, referencias, etc. Todos estos elementos aparecen declarados en una DTD específica para cada uso (Archivo e Intercambio, Edición, Autoría). Ampliaremos estos aspectos cuando tratemos la naturaleza de JATS en el capítulo 3, sección 3.

XML está siendo utilizado más que simplemente para intercambiar ficheros o para generar impresos, PDF, HTML u otros formatos de salida. En el caso de los contenidos de revista, la adopción de XML con un estándar de marcado como JATS permite la extracción y enriquecimiento de contenido semántico útil, como denominaciones de compuestos químicos y biológicos, nombres de personas como autores y editores, nombres de lugares, de organizaciones, direcciones, fuentes de financiación y las relaciones entre estos elementos.

Dentro de las ventajas más relevantes de la codificación basada en XML para los lenguajes de marcado podemos mencionar: los lenguajes basados XML pueden ser utilizados en cualquier plataforma, pues los archivos XML son archivos de texto y admiten distintas codificaciones de caracteres (ISO 8859, UTF-8, entre otros); ofrecen además una excelente interoperabilidad entre sistemas al ser XML un estándar abierto, no propietario, desarrollado por W3C, un consorcio

internacional sin fines de lucro; los elementos básicos definidos en forma de árbol brindan a los lenguajes basados en XML simplicidad y usabilidad; los esquemas son extensibles: un esquema codificado con XML puede ser diseñado e implementado teniendo en cuenta el futuro crecimiento y expansión del mismo (si se presenta la necesidad de describir nuevos datos, el esquema existente puede ser modificado); son adaptables: un esquema desarrollado en XML es adaptable no solo a los requerimientos futuros de los documentos sino también al crecimiento de los sistemas en cuanto a funcionalidad.

Otra ventaja muy apreciada de los documentos marcados con esquemas basados en XML es que la representación de los documentos XML en un formato de salida determinado (presentación), se define generalmente por medio de hojas de estilo. Éstas son las que permiten la separación de la estructura de la presentación del contenido del documento, y posibilitan procesar la presentación por separado. Por ejemplo, un navegador puede procesar la información contenida en los elementos y atributos, y presentarlos en HTML. También se pueden obtener otras representaciones de ese mismo contenido en formatos de archivo PDF o EPUB. Tanto la versión HTML como la PDF o la EPUB de este ejemplo, son generadas a partir del mismo documento XML, pero utilizando diferentes hojas de estilo.

El uso que se le da a los documentos XML incluye la localización, recuperación, presentación, interpretación, y reutilización de los datos y la información en ellos contenida y sus metadatos asociados. El acceso a estos datos es realizado tanto por humanos como por sistemas de software, dentro o fuera del contexto en que los datos o los documentos fueron creados. Comparado con las bases de datos y los repositorios de documentos tradicionales, los repositorios de documentos XML son más complejos en muchos aspectos. Por ejemplo, en sus contenidos se utilizan lenguajes formales y naturales, las estructuras y el marcado de los documentos en estos repositorios es a veces heterogéneo, muchas veces las necesidades de acceso conciernen a unidades de contenido con diferentes niveles de granularidad, y los criterios de acceso pueden estar relacionados con la estructura, con el contenido o con el contexto de la información (Salminen, Jauhiainen y Nurmekselä, 2014).

Las revistas electrónicas tienen muchas ventajas sobre las impresas: incluyen enlaces en las referencias, enlaces a acciones concretas, y capacidades de búsqueda sofisticadas. Sin embargo, las revistas electrónicas no son un reemplazo aceptable para la versión impresa sin la seguridad de que el contenido podrá ser preservado de forma que se asegure la recuperación de los archivos en el futuro. Los esquemas para el contenido de revistas electrónicas definidos en XML proporcionan un mecanismo adecuado para la preservación y recuperación futura de su contenido a largo plazo (Abrams y Rosenblum, 2003).

Además de preservar la estructura del artículo, un fichero XML puede ser interpretado por humanos y por máquinas. Un esquema de marcado para el contenido de los artículos escrito en un lenguaje abierto, que es además un estándar de W3C, extensible, adaptable y multiplataforma, se presenta como un excelente candidato cuando se trata de asegurar la preservación y la recuperación a largo plazo de los contenidos de las revistas electrónicas.

*In order for the use of e-journals to fulfill their promise to enhance the scholarly literature and modes of discourse, it is necessary to provide an archival form for e-journal content that is acceptable to content providers, consumers, and archival institutions. A domain neutral schema for e-journal content in a manner that is amenable to long-term preservation and retrieval, and can thus help in facilitating an online-only future for the scholarly literature. Such a schema should be independent of academic discipline, journal format, and the proprietary policies and practices of publishers and archiving institutions (Abrams y Rosenblum, 2003).*



La granularidad con la que cada artículo debe ser marcado ha sido motivo de debate. Murray-Rust y Rzepa (2003) proponen marcar cada aspecto de un documento científico, desde números, tablas, gráficos, diagramas, secuencias, imágenes ("*micro-markup*"). Lo que ellos proponían era un cambio en la forma de publicación tradicional, y sostenían que "*XML can revolutionise the way that scientists make their work available*". Otros como Brown (2003), más enfocado en las ventajas del marcado desde el punto de vista comercial, defendía el marcado estructurado como forma de mejorar la publicación, con una postura a favor de un modelo que representase correctamente el contenido en diferentes medios y que tuviese suficiente marcado semántico para su posterior utilización en recursos en línea.

La granularidad que se pretenda alcanzar afecta las decisiones respecto al diseño de un esquema, el cual debe ser coherente con su propósito. Por ejemplo, en lo que respecta al nivel de profundidad que se quiere alcanzar en el marcado de un documento XML, y tomando el caso de las referencias bibliográficas, éstas pueden marcarse con etiquetas de comienzo y final <REF> ... </REF> (baja granularidad) o puede marcarse cada elemento de la referencia: autores, fecha, título del artículo, volumen de la revista, número de la revista, paginación (alta granularidad). Continuando con este ejemplo de marcado de las referencias, para la versión impresa del artículo la sección de referencias podría comenzar con <REF> y las referencias ser tratadas como párrafos con un patrón particular de negrita, cursiva y puntuación. En cambio, para aprovechar toda la potencialidad de un artículo en el mundo digital, los ordenadores deben poder distinguir el autor principal, el título abreviado de la revista, su volumen y número, la fecha de publicación, etc., en cada una de las referencias. Es por ello que en el entorno electrónico generalmente se requiere un marcado más granular que para la edición impresa.

*In their document analysis process, publishers need to make decisions about what level of granularity is necessary and practical for their content and modes of publication. It is also wise to consider at what stage of the editorial and production process it makes the most sense to do the more detailed tagging* (Kasdorf, 2003).

Sucede que muchas veces para generar metadatos de calidad, los artículos y sus elementos son marcados de forma granular utilizando semántica y taxonomías personalizadas, o ambos. A veces el marcado es manual, a veces automatizado, y a veces un poco de ambos. Pero todo el proceso debe ser pensado y adecuado a sus propósitos, incluyendo en las decisiones de diseño previsiones como que luego debe ser mantenido, mejorado, ampliado, a veces migrado o corregido y todo tiene un costo asociado.

## CAPÍTULO 3. ANTECEDENTES, ORIGEN, EVOLUCIÓN Y NATURALEZA DE JATS

En este capítulo se presentan los antecedentes relacionados con lenguajes de marcado respecto a su utilización para el marcado de contenidos de revistas científicas, y se describe la evolución de estos lenguajes. Interesan particularmente aquellos que de una u otra manera son antecedentes directos de JATS, a los cuales se dedica la segunda sección de este capítulo.

JATS es producto de la experiencia en desarrollo de esquemas de marcado para *PubMed Central* (PMC). Se origina como resultado de la evolución de esquemas anteriores, los cuales son fruto de la colaboración entre la comunidad de expertos y usuarios de PMC, que se vinculan y trabajan juntos desde hace varios años. Estudiar los lenguajes antecesores de JATS es conocer cómo las comunidades de desarrolladores involucradas han ganado experiencia para llegar a crear el estándar actual.

[...] *JATS has proven to the community that standardization brings cost savings and scalability. Twenty years ago, the community didn't have enough operational experience to successfully create a standard tag set* (Rosenblum, 2015).

La tercera y última sección del capítulo trata sobre la naturaleza del estándar, mediante una introducción general acerca de sus cometidos y aplicación, para pasar a presentar luego las ventajas y desventajas de utilizar un lenguaje estándar de marcado como JATS.

### 3.1. Antecedentes: normas y estándares para el marcado de revistas

Rosenblum y Golfman (2002) sitúan los antecedentes de las DTD actuales para la estructuración trabajos académicos en el trabajo que la *Association of American Publishers* (AAP) realizó desde 1983 a 1987, cuya experiencia fue utilizada en la elaboración de la norma ANSI/NISO Z39.59 de 1988, la cual a su vez sirvió de base para el estándar ISO 12083 y su *Serial DTD*. Todos estas DTD están desarrolladas para SGML, el metalenguaje existente utilizado en esa época.

SGML es un metalenguaje, normalizado por la ISO 8879:1986, y fue desarrollado sobre la base de GML, un metalenguaje que se creó en 1969. Con esquemas basados en SGML se buscaba un lenguaje que permitiese a las editoriales separar las instrucciones de formato (que suelen ser lenguajes propietarios) de la información estructural, mediante el etiquetado de contenidos por su significado semántico en lugar de por el formato de presentación. El formato de presentación se puede aplicar a los elementos estructurales con una hoja de estilo cuando se necesita mostrar el contenido (Rosenblum y Golfman, 2002).

En diciembre de 1989, se forma el grupo de trabajo conocido como *European Workgroup for SGML* (EWS), liderado por tres de los mayores editores de revistas, *Elsevier Science*, *Wolter Kluwer Academic Publishers* y *Springer Verlag*. El objetivo de ese grupo era desarrollar una DTD estándar para los encabezados de los artículos de revista, y para ello se basaron en los resultados del proyecto de AAP. Es así que en enero de 1991 se publica MAJOUR, "*Modular Application for JOURnals*", la primer DTD desarrollada para marcar metadatos en artículos de revista. El motivo principal de su desarrollo fue el requerimiento de metadatos para un cada vez más creciente y diverso panorama en cuanto a recuperación de información en diferentes bases de datos en línea (Cave, 2003).

En 1994 se publica el estándar ISO 12083. Este estándar ISO proporciona estructuras de documentos para la creación y el intercambio de libros, artículos y publicaciones seriadas, especificados en cuatro tipos de DTD, correspondientes a los siguientes tipos documentales: libros, publicaciones periódicas, artículos, y una definición de tipo de documento para matemáticas que puede ser integrada en otras aplicaciones SGML (ISO, 1994).

Otro intento por definir un estándar que lograra amplia adopción fue el que protagonizaron varias editoriales de revistas académicas de Reino Unido en 1995 cuando formaron el grupo de trabajo *ad hoc* OASIS (*Organization for Article Standards In Science*). Este grupo trabajó en el desarrollo de SSSH: *Simplified SGML for Serial Headers*. La DTD de SSSH se basó en MAJOUR, aunque no son enteramente compatibles. La primera versión de SSSH DTD fue publicada en 1996 y la segunda versión, SSSH2 DTD, en 1998 (Cave, 2003).

En los 90, SGML era utilizado por industrias y organizaciones con altos requerimientos de intercambio: la industria aeronáutica, la automovilística, y la defensa. Pero la edición académica en esos años era un ámbito muy fragmentado y por esa razón la ISO 12083 no logró consolidarse como un estándar ampliamente aceptado para las publicaciones científicas académicas. En ese momento se buscaba un estándar que posibilitara múltiples formatos de salida, y un lenguaje de marcado que sirviera tanto para generar versiones impresas como en línea (Owens, 2010).

Si bien la norma ISO 12083 y su conjunto de DTD no tuvieron éxito en su objetivo de convertirse en un estándar ampliamente adoptado por las editoriales, ni para libros ni para revistas, se constituye sin dudas en un antecedente relevante para las bases estructurales de las DTD que se han creado posteriormente. Según Rosenblum y Golfman (2002), ese objetivo de convertirse en estándar de la industria editorial no se alcanzó debido a que, si bien las DTD tomaban en cuenta un panorama general de las necesidades de la industria, las mismas no cumplían con los requisitos de negocio que eran específicos de cada organización dentro de cada comunidad dedicada a la publicación académica.

Muchas editoriales optaron por desarrollar sus propias DTD con SGML utilizando ISO 12083 o MAJOUR como base, pero pocas utilizaron estos estándares sin agregarle modificaciones propias (Wusteman, 2003b).

### 3.2. Origen y evolución de JATS

Los orígenes del modelo para artículos creado por *PubMed Central* (PMC) se remontan a 1999, cuando la especificación de ese lenguaje de marcado se recogía exclusivamente en DTD. *pmc-1.dtd* sirvió de base para el posterior desarrollo de *NLM DTD Suite* en 2003. Lapeyre (2008) hace una muy buena síntesis de la filosofía que dio origen al modelo NLM DTD: “*A rose is a rose is a rose. I can’t tell you what to publish. We may not model everything exactly the same way, but if we both publish the same logical or semantic structure (paragraph, table, figure, author name, journal abbreviated name) we name it (tag it) the same thing.*”

Si bien podemos rastrear los orígenes del marcado de artículos a la *APP Serial DTD*, los antecedentes directos de JATS están en los desarrollos llevados a cabo por la *National Library of Medicine* (NLM) en el marco de su proyecto PMC. Los esquemas que ha desarrollado PMC y son considerados antecesores de JATS, se han agrupado en tres etapas (Figura 1), indicando la fecha de publicación de cada versión entre paréntesis. La primera etapa es la que hemos denominado orígenes e incluye las dos versiones existentes de NLM DTD. En la segunda etapa hemos incluido todas las versiones de *NLM Tag Suite*. La tercera etapa agrupa el diseño actual del estándar e incluye la evolución de sus versiones hasta llegar a la actual: JATS v. 1.1 de noviembre de 2015.



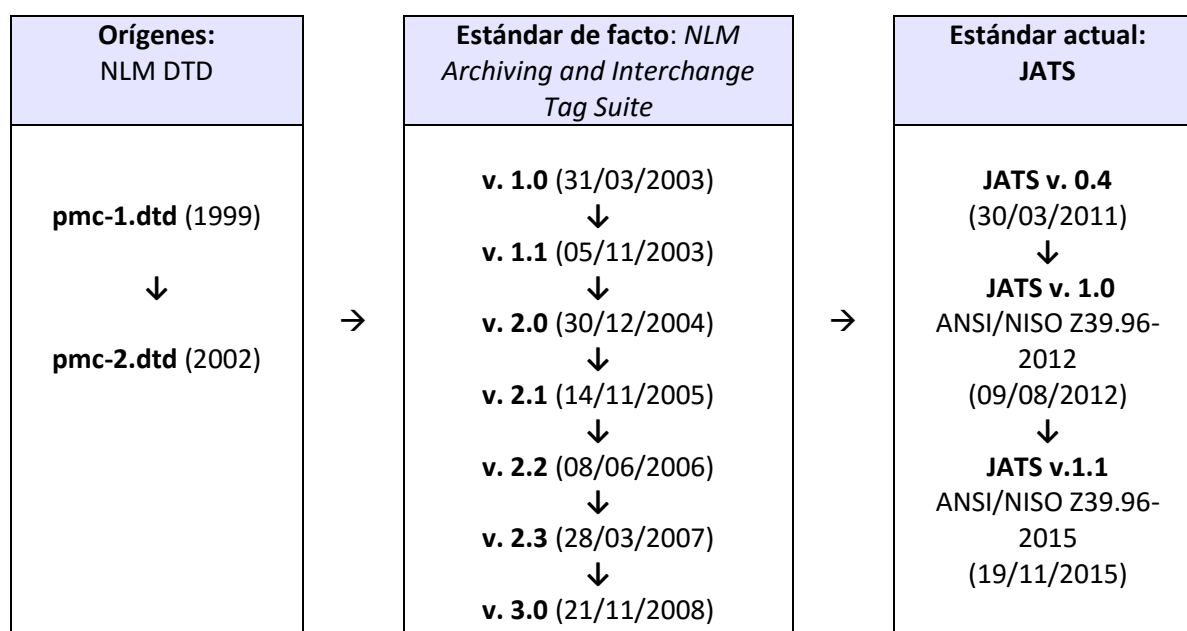


Figura 1. Evolución de JATS  
Fuente: elaboración propia.

### 3.2.1. Orígenes. Las primeras versiones de NLM DTD

El requerimiento técnico fundamental para los editores en los inicios de PMC en 1999 era proveer los artículos en formato SGML o XML. La necesidad de unificar modelos de contenido en un solo modelo normalizado para los artículos se debió a que simplificaba la ingesta y la carga de contenidos en el software encargado de hacer accesibles los artículos en la web.

Beck y Usdien Tommie (2013) explican que la primer DTD de *PubMed Central*, pmc-1.dtd, fue escrita en base a dos modelos de contenido que correspondían a los artículos que se publicaban en PMC en 1999. Según Beck (2010), estas dos DTD de base para la creación de pmc-1 fueron keton.dtd, una DTD en SGML descendiente de la ISO 12083 y la DTD de *BioMed Central* denominada article.dtd, una DTD en XML.

En 2001, la Biblioteca de la Universidad de Harvard, preocupada por el archivo, almacenamiento y recuperación a largo plazo de la información contenida en las revistas electrónicas, encargó como parte de su *E-journal Achiving Project*, un estudio sobre la viabilidad de una única DTD que pudiera ser utilizada para archivar todas las revistas electrónicas de su fondo. La empresa consultora encargada de realizar el estudio, Inera Inc., elaboró un informe en base a la información recabada respecto a las DTD utilizadas por las siguientes editoriales: *American Institute of Physics*, *BioOne*, *Blackwell Science*, *Elsevier Science*, *Highwire Press*, *Institute of Electrical and Electronics Engineers*, *Nature Publishing Group*, *PubMed Central*, *University of Chicago Press*, *John Wiley y Sons*. El informe llegó a la conclusión que una única DTD adaptable a cualquier artículo de revista electrónica era posible, pero que ninguna de las DTD analizadas en el estudio cumplía con todos los requisitos. Dentro de las recomendaciones de este informe sobre el diseño y desarrollo de una nueva DTD se destacan la recomendación de uso de XML y no SGML para el archivo del contenido, fundamentado en la existencia de una mayor cantidad de herramientas para XML. Recomiendan además que la DTD

para archivo deberá ser menos restrictiva en estructura y más eficiente en la selección de elementos que las DTD específicas creadas por las editoriales, ya que debe adaptarse a una amplia gama de estilos de revista (Inera, 2001).

En 2002 NLM, Harvard Library, Mulberry Technologies e Inera comienzan a trabajar de forma conjunta para modificar pmc-1.dtd basándose en las recomendaciones del informe sobre el estudio de viabilidad encargado por Harvard en 2001.

Este grupo de trabajo sentó las bases de lo que sería una nueva DTD, acordando que se trabajaría en un conjunto de elementos y atributos “estándar” en XML, que pudiera ser utilizado para construir modelos para los artículos. Estuvieron de acuerdo en continuar trabajando en el nuevo modelo, para ampliarlo y que pudiera servir para marcar cualquier contenido en los artículos, e incluyeron en su planificación de trabajo la realización de una encuesta que comprendiera varias disciplinas, con el objetivo de asegurarse que todos los objetos pudieran ser acomodados en el nuevo modelo. Además acordaron que debía haber dos modelos para los artículos, uno para el contenido ya existente, que permitiera convertir datos de múltiples fuentes a la nueva DTD y otro para la creación de nuevo contenido, más prescriptivo y con reglas explícitas de cómo marcar el contenido (Beck y Randall, 2013). Esos fueron los orígenes de pmc-2.dtd, una DTD de uso general para cualquier artículo de revista en formato electrónico.

### 3.2.2 NLM *Journal Archive and Interchange Tag Suite* como estándar de facto

La primera versión de *NLM Journal Archive and Interchange Tag Suite* (también conocida como NLM DTD), la cual podemos considerar como el antecedente más cercano de JATS, fue publicada a principios de 2003, y se convirtió rápidamente en el estándar *de facto* para marcar artículos de revista con XML (D.A. Lapeyre, 2009) (Jeffrey Beck, 2011).

*Since the introduction of the first “NLM DTD” in 2003, the set of DTDs developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM) has steadily gained acceptance as a quasi-standard for using XML markup to produce, exchange, and archive journal articles and books within the scientific, technical, and medical (“STM”) publishing spaces (O’Brien, 2011).*

El Conjunto de etiquetas NLM para archivo e intercambio de revistas (*NLM Journal Archive and Interchange Tag Suite*) tuvo actualizaciones periódicas con el asesoramiento del Grupo de Trabajo que se denominó *NLM Archiving and Interchange Tag Suite Working Group*. Ese grupo fue el responsable de las 7 versiones existentes del conjunto de etiquetas.

La versión 1.0 contaba con dos conjuntos de etiquetas: Archivo e Intercambio - *Archiving and Interchange (Green)* y Publicación de revistas - *Journal Publishing (Blue)*.

Según el sitio web oficial de NLM dedicado a la documentación de las versiones, en la versión 1.0, la DTD para archivo e intercambio es presentada como un conjunto de módulos XML que nombran y definen los elementos y sus atributos, con la finalidad de describir el contenido textual y gráfico de artículos de revistas (incluyendo también algunos materiales que no son propiamente artículos, tales como cartas, editoriales y comentarios de libros). Los módulos fueron desarrollados como parte de un esfuerzo para crear aplicaciones XML a través de las cuales los materiales en disciplinas relacionadas con la salud puedan ser compartidos y reutilizados electrónicamente. Los módulos se pueden utilizar para construir DTD para describir artículos y también para archivarlos, además de poder construir DTD para la transferencia de artículos de revistas desde las editoriales a los archivos y entre archivos. Aunque la *Suite* completa fue desarrollada para apoyar la producción electrónica, las estructuras son también adecuadas para la producción de versiones impresas.

La DTD de publicación de revistas de la versión 1.0 admite tanto describir los metadatos de un artículo de revista y el contenido del artículo, como también describir solamente los metadatos de la cabecera del artículo. Es una DTD de tipo prescriptivo, optimizada para la creación y el etiquetado XML de contribuciones de revistas. Aunque fue diseñada para revistas biomédicas, la DTD está pensada para ser lo suficientemente general para describir no sólo las revistas del área ciencia, tecnología y medicina (STM) sino además otras revistas en cualquier campo. Los principios de diseño de las DTD se tratarán con más detalle en el capítulo 4.

La versión 1.1 de noviembre de 2003 sigue manteniendo los dos conjuntos de etiquetas de la versión 1.0, pero amplía la cantidad de elementos. La principal novedad de esta versión es que además de las DTD, incorpora esquemas XML del W3C para cada conjunto de etiquetas.

La versión 2.0 de 2004 introduce la división en tres conjuntos de etiquetas en vez de dos: un conjunto orientado a la preservación, otro orientado a la publicación y un tercer conjunto menos extenso y más estricto orientado a la creación de artículos. Otra diferencia importante con la versión anterior es que se realizó una remodelarización del conjunto de etiquetas para satisfacer nuevos requisitos y con el propósito de hacer más evidente cómo utilizar el nuevo conjunto. El conjunto de etiquetas de *National Center for Biotechnology Information* (NCBI) para libros (*NCBI Book Tag Set*) también fue introducido con la versión 2.0. NCBI para libros fue creado con la intención de proporcionar un formato común para la creación de contenidos de libros de texto en XML.

Los principales cambios de la versión 2.1 de 2005 fueron originados en la necesidad de adaptación a las nuevas versiones del esquema MathML de W3C, lo cual derivó en la necesidad de modificar los módulos del conjunto de etiquetas NLM para reflejar la nueva MathML. La versión 2.1 incluyó por primera vez el conjunto de etiquetas en RelaxNG.

La versión 2.2 de 2006 fue una mejora de la versión anterior, basada en cambios solicitados por NLM y otros usuarios, en la que se realizaron solo cambios menores relacionados con los valores de algunos atributos. Lo mismo se puede decir de la versión 2.3, que mejoró la versión precedente, esta vez a solicitud del Grupo de trabajo, fundamentada en cuestiones operacionales que debían ser solucionadas.

Los cambios en la versión 3.0 fueron muchos, lo que le da a esta versión la particularidad de no ser compatible con ninguna de las versiones anteriores. Algunos cambios implicaron modificar el nombre de elementos para que la denominación fuese más coherente en todo el conjunto de etiquetas, o la adición de algunos elementos que pasan a cumplir la función de agrupar otros elementos ya existentes. Cambios como éstos son los que afectaron a la compatibilidad con versiones anteriores.

Beck y Usdien Tommie (2013) enfatizan que las modificaciones están basadas en los requerimientos de usuarios reales y en habilitar lo que los editores ya estén haciendo con sus contenidos, no en tratar de definir qué es lo que deberían hacer, o basadas en predicciones de lo que se podría necesitar en el futuro. Según ellos, tanto los Grupos de Trabajo de NLM como los de NISO JATS centraron sus roles en la normalización y la documentación de las prácticas ya existentes para ayudar en el uso, reutilización e intercambio de contenido actual y futuro de los artículos, pero no en tratar de influir en el futuro de la edición de artículos.

### 3.2.3 Estándar JATS

Cuando el uso de *NLM Journal Archive and Interchange Tag Suite* comienza a crecer, son los usuarios y potenciales usuarios quienes comienzan a demandar la formalización del este modelo de marcado de *PubMed Central* en un estándar. El trabajo conjunto de NLM y NISO comenzó a finales de

2009 y el primer borrador de JATS (v. 0.4) para pruebas de uso fue publicado en marzo de 2011 como norma NISO Z39.96 (Beck y Usdien Tommie, 2013).

Se hizo necesario estandarizar un lenguaje de marcado y la forma de trabajar con ese lenguaje en la publicación de revistas, pues se había convertido en un lenguaje utilizado por muchos, extendiéndose su uso mucho más allá del proyecto original para el que fue creado el conjunto de etiquetas. Antes de que JATS se convirtiese en estándar, la evolución del conjunto de DTD había estado estrechamente relacionada con las necesidades y los requerimientos de *PubMed Central*, el portal de acceso a revistas biomédicas la *National Library of Medicine* de Estados Unidos. A partir de que se comienza a trabajar en pro de un estándar, la propia conformación del nuevo grupo de trabajo NISO se vuelve más heterogénea. Si bien muchos de sus integrantes provienen del Grupo de trabajo de NLM (integrado por representantes de NCBI de la NLM, representantes de Mulberry Technologies y de Inera), se suman además integrantes provenientes de otras instituciones/organizaciones, tanto públicas como privadas: *Microsoft*, *ITHAKA/JSTOR/Portico*, *American Institute of Physics*, *HighWire Press*, *OHIO Link* y *Atypon Systems Inc.*

*NISO JATS: Journal Article Tag Suite, version 1.1* es la denominación actual del lenguaje de marcado estandarizado para artículos de revistas, también conocido como estándar **ANSI/NISO Z39.96-2015**, aprobada por *American National Standards Institute* (ANSI) el 19 de noviembre de 2015. Actualiza la versión anterior del estándar publicada en julio de 2012. Además de las descripciones de elementos y atributos, proporciona los tres conjuntos de etiquetas (Archivo e intercambio, Edición y Autoría) como parte de NISO JATS 1.1 (*JATS: Journal Article Tag Suite, version 1.1*, 2015).

Desde que la primera versión fue publicada (JATS v.1.0), ha habido tres borradores en desarrollo, todas versiones no normativas en las que ha trabajado el Comité Permanente (*JATS Standing Committee*): 1.1d1 (el primer borrador de la versión 1.1 de diciembre de 2013), 1.1d2 (la segunda versión borrador de diciembre de 2014) y 1.1d3 (tercera versión borrador de trabajo, de abril de 2015). La versión actual (ANSI NISO Z39.96-2015) incluye todos los cambios realizados en base a los comentarios del público sobre las tres versiones del Comité de trabajo del estándar. Los comentarios sobre JATS 1.0 realizadas a través del formulario que proporcionó NISO en noviembre de 2015 también han sido considerados por el Comité Permanente de JATS e incorporados en esta última actualización. Todos los cambios aceptados son compatibles con JATS 1.0, lo que significa que cualquier documento que era válido para JATS 1.0 será válido para JATS 1.1.

Los cambios en la *Tag Suite* se deben a las necesidades detectadas por los usuarios. Los cambios sugeridos por los usuarios son recogidos y analizados por el Comité Permanente de JATS, quien decide sobre las acciones a tomar respecto a las propuestas recibidas en base a los procedimientos de mantenimiento continuo establecidos por NISO.

Un ejemplo de los cambios introducidos y aprobados, y que forman parte de las adaptaciones más recientes introducidas a la última versión oficial del estándar, es la propuesta para incorporar citas de conjuntos de datos (*data sets*), elevada al *JATS Standing Committee* por parte del grupo de trabajo *JATS Data Citation Working Group*, el cual planteó la incorporación de esta adaptación a la versión borrador 1.1d2 de JATS.

*JATS – like its precursors – has had provisions for handling citations from the start. The current standard NISO JATS 1.0 (henceforth JATS 1.0) as well as the draft standard JATS 1.1d1, however, are limited in how they can handle data citations* (Mietchen y otros, 2015).

El depósito de conjuntos de datos en repositorios, y la citación de dichos datos en los artículos correspondientes, es y ha sido una práctica común en ciencias biológicas y naturales. También lo es vincular esos conjuntos de datos con los artículos y muchas comunidades de investigación, revistas y

repositorios colaboran para que ello sea posible. Desde que las revistas se hicieron predominantemente accesibles en línea, los ficheros conteniendo conjuntos de datos suplementarios están vinculados mediante enlaces en el propio artículo.

Evidenciada la necesidad, el desafío planteado fue cómo adaptar esta infraestructura de marcado para dar cabida a las necesidades de la citación de datos. Se consideraron dos alternativas para incluir la opción de citar conjuntos de datos: la modificación del estándar JATS o la creación de una extensión del estándar JATS. Las dos alternativas se discutieron en una reunión llevada a cabo en junio de 2014 en Londres, por un grupo integrado por 18 editores y expertos en JATS (*The Force 11 JATS Data Citation Working Group*), y se acordó rápidamente que la infraestructura existente en JATS para marcar citas ya era lo suficientemente comprehensiva como para que con unos pocos cambios se pudieran contemplar las necesidades de citación para conjuntos de datos identificadas por el grupo. El beneficio de este enfoque es que no fue necesario repetir los elementos requeridos, pues muchos de los elementos necesarios para citas de conjuntos de datos son los mismos que se utilizan para las citas de artículos (Mietchen y otros, 2015).

### 3.3. Naturaleza de JATS

JATS proporciona un formato XML común en el que las editoriales y los archivos pueden intercambiar contenidos de revista. El esquema JATS ofrece un conjunto de elementos y atributos XML para describir el contenido textual y gráfico de artículos de revista, así como algunos materiales que no son artículos y también se publican en las revistas: cartas, editoriales, y reseñas de libros o de productos (National Information Standards Organization, 2012).

*The Journal Article Tag Suite (JATS) is a description of a set of elements and attributes that is used to build XML models of journal articles for archiving, publishing, and authoring* (Beck y Usdien Tommie, 2013).

El conjunto de etiquetas JATS define los elementos y atributos que describen tanto los metadatos como el contenido completo de los artículos de revistas científicas y académicas. El conjunto de etiquetas no está diseñado para describir revistas de divulgación, libros, u otros formatos de publicación. En el propio texto del estándar se advierte que a pesar de que algunas estructuras en estos tipos de documentos antes mencionados pueden ser similares a las estructuras de los artículos de revistas, otras pueden ser muy diferentes y no podrán ser tratadas de forma adecuada con los elementos y atributos existentes en el conjunto de etiquetas JATS.

Este estándar define un conjunto de estructuras (elementos y atributos) que se pueden utilizar para crear una variedad de conjuntos de etiquetas para usos específicos. El documento del estándar describe el conjunto completo elementos y atributos predefinidos, denominado "*Tag Suite*", los cuales son la base para crear conjuntos de etiquetas. Junto a esas descripciones, se incluyen los tres modelos o conjuntos de etiquetas denominados "*Tag sets*": *The Journal Archive and Interchange Tag Set*, *The Journal Publishing Tag Set* y *The Article Authoring Tag Set*. Son los mismos tres modelos base de *NLM Tag Suite* v. 3.0, motivo por el cual la versión 1.1 de JATS es totalmente compatible con NLM versión 3.0.

Cada uno de los *Tag sets* o conjuntos de etiquetas de JATS son esquemas contruidos en base a los componentes descriptos en el *Tag Suite*, creados con un cometido específico.

El conjunto de etiquetas de archivo e intercambio - *Archiving y Interchange (Green) Tag Set* -es un conjunto de etiquetas de base para repositorios XML, posibilita la conversión desde otros conjuntos de etiquetas pues captura convenientemente la estructura y la semántica, proporcionando un formato común para el intercambio XML entre editores, archivos, agregadores y proveedores de servicios



(Beck, J., 2010). Además, permite preservar los metadatos y el contenido del artículo, incluyendo la preservación de números, símbolos y puntuación en las referencias bibliográficas.

El conjunto de etiquetas de publicación - *Journal Publishing (Blue) Tag Set* - es un modelo más prescriptivo que el anterior. Es un modelo común para facilitar la conversión del contenido de las revistas a XML y para publicar partiendo de documentos en XML. Está diseñado para proporcionar un formato de datos organizado de forma tal que facilite la producción de artículos tanto en la web como impresos. Para lograr este objetivo, el conjunto de etiquetas de publicación establece secuencias de elementos más estrictas y limita las opciones para el etiquetado.

El conjunto de etiquetas de autoría - *Article Authoring (Pumpkin) Tag Set* - es el modelo más prescriptivo de los tres. El contenido en XML puede ser escrito por el autor directamente. El control de contenido y la secuencia de los elementos se aplican estrictamente. Es un modelo centrado en la creación del artículo por parte de sus autores y no contiene metadatos referentes al número o a la revista.

Los *Tag Sets* son modelos que agrupan de forma específica elementos y atributos definidos en la *Tag Suite*. Cada uno de los tres *Tag Sets* tiene un color asignado (para poder referirse a ellos como conjunto de etiquetas verde, azul o naranja), describe un modelo estructural específico y emplea una cantidad de elementos y atributos que varía según el *Tag Set* (Tabla 1).

Tabla 1. Cantidad de elementos y atributos empleados en cada *Tag Set*, JATS v.1.0

Fuente: elaboración propia.

Denominación del <i>Tag Set</i>	Cantidad de elementos	Cantidad de Atributos
Archivo e intercambio (verde) / <i>Archiving y Interchange (Green)</i>	254	135
Publicación de revistas (azul) / <i>Journal Publishing (Blue)</i>	246	134
Autoría de artículos (naranja) / <i>Article Authoring (Pumpkin)</i>	197	123

La *Suite* de etiquetas JATS es en general descriptiva, más que prescriptiva. Fue diseñada para permitir la conversión desde otros modelos de marcado. Es por ello que muchos de sus elementos son opcionales. Se trata de preservar la mayor cantidad posible de semántica en el contenido XML que provenga de otros conjuntos de etiquetas. Esto significa que fue concebido para describir casi todo lo que se pueda encontrar en los contenidos de cualquier editorial. De esta manera, la información etiquetada diferente por diversas editoriales no se pierde cuando se convierte a JATS.

Los conjuntos de etiquetas JATS han sido diseñados para ser extensibles. Cualquiera de los conjuntos de etiquetas puede ampliarse o restringirse para satisfacer las necesidades de un determinado proyecto. Existe también la posibilidad de crear nuevos conjuntos de etiquetas a partir de los elementos y atributos de JATS.

Otra característica de JATS, que hereda de su antecesor NLM DTD es que contempla etiquetas para diferentes tipos de contribuciones a revistas, no solamente para los típicos artículos de investigación original o artículos de revisión.

Los esquemas correspondientes a los 3 conjuntos de etiquetas están disponibles en DTD, W3C Schema y RelaxNG. Estos esquemas, así como la información detallada del uso de cada uno de sus conjuntos de etiquetas no se incluyen en el texto del estándar, sino que están accesibles desde el espacio que NLM le ha dedicado a JATS en su sitio web. El documento del estándar tampoco incluye otra información no normativa como las librerías de etiquetas (*Tag Libraries*), ni las recomendaciones de uso. Para consultar esa información “no normativa” se debe acceder al sitio de NLM dedicado a JATS e ingresar a la sección *Non-normative supporting information*.

La apuesta de NLM y NISO por definir JATS a nivel de estándar nacional se fundamenta en varias ventajas, como las mencionadas por Beck y Randall (2013): *“Normalizing the structure of journal articles enables interchange of articles among publishers, authors, data conversion vendors, and aggregators such as archives and indexing services. An existing, well used, and freely available article model also allows new, small journal publishers to start creating articles in XML significantly faster, more easily, and at less cost than if they had to create a model and persuade their vendors and publishing partners to use it.”* La utilización de un estándar de marcado facilita el intercambio de artículos entre diferentes editores, autores, y agregadores de contenido como los repositorios o los servicios de indización. Una estructura normalizada facilita la conversión de contenido y el intercambio de datos.

En opinión de Carpenter (2014), *“As digital publishing evolves, the importance of common vocabulary structures like JATS will only increase, because exchanging digital files is a critical component in a functioning digital content ecosystem.”*

Dentro de las ventajas de implementar JATS en la actualidad se encuentra el hecho de que ya existen XSLT/XSL:FO para las transformaciones, y también hay herramientas varias de conversión, así como también manuales con la metodología de marcado. Además, es más fácil encontrar software disponible que pueda procesar archivos XML marcados con un estándar. Los analizadores XML o *parsers* (aplicaciones de software) pueden utilizar una DTD estándar para comprobar que los documentos marcados cumplen con las especificaciones definidas en ella.

Las desventajas de adoptar un lenguaje estándar de marcado están relacionadas con la forma en que éste se desarrolla y actualiza, pues lógicamente estos procesos no están supeditados a necesidades específicas de una editorial en particular, y si se quiere personalizar el contenido del esquema es difícil hacerlo sin apartarse de la norma. Esos problemas se experimentan siempre que se prefiere adoptar un estándar. Además, de realizarse modificaciones en el esquema, hay otras consecuencias de tipo instrumental, como tener que alterar los flujos de trabajo editorial, o que el software diseñado para procesar el contenido con ese estándar no funcione como debería. En esos casos se debe ser muy consciente de las implicaciones que puede tener aplicar variaciones, por mínimas que sean, si se decide implementarlas. Además, se debe tener en cuenta que cuando los estándares se adaptan dando lugar a personalizaciones del estándar original, éstas pueden o no ser interoperables a largo plazo, incluso si son compatibles hacia atrás en el corto plazo. Esto presenta un problema a la preservación digital del contenido, como se verá más adelante en el capítulo 6 para el caso de algunas personalizaciones de JATS.

## CAPÍTULO 4. LÍNEAS DE DISEÑO Y FUNCIONALIDADES

Existen diferentes aproximaciones al diseñar un esquema de marcado para documentos basado en XML. El estándar JATS, tal como se ha descrito en el capítulo anterior, está compuesto por una *Tag Suite*, o conjunto de definiciones estructurales formadas de elementos y atributos XML (de ahora en más *Suite*), y tres conjuntos de etiquetas o *Tag Sets*<sup>1</sup>, creados para usos específicos. Esencialmente, la *Suite* define los componentes y los *Tag Sets* son los modelos descriptivos para el contenido de los artículos. Las estructuras de los tres *Tag Sets* comprendidos en el estándar están diseñadas de acuerdo a la función principal para la cual se describe el contenido del artículo. La presencia y el orden de los componentes estructurales preestablecidos varían según el modelo, pudiendo ser más permisivo o restrictivo un *Tag Set* de acuerdo a su función.

En este capítulo se presentan las principales líneas de diseño de JATS y sus tres modelos específicos, así como también la forma de estructurar el marcado por parte de cada uno de los tres *Tag Sets*, mediante la comparación de los diagramas de jerarquía disponibles en las respectivas librerías de etiquetas (*Tag Libraries*).

### 4.1. JATS: principales líneas de diseño

Para diseñar un esquema de marcado en XML, se parte de analizar el tipo de documento objeto de la descripción, con vistas a seleccionar qué elementos del contenido deben ser definidos, decidir cómo nombrarlos, establecer cómo se relacionan entre ellos, reconocer y codificar sus variantes (atributos) y determinar qué información no textual acerca del contenido puede ser útil. El resultado de analizar las características de los contenidos de revista es un esquema (*document model* en la literatura en inglés) al que nos referiremos de aquí en adelante como **modelo de documento** o simplemente **modelo**. Ese modelo necesita ser formalizado en un lenguaje que permita a humanos y ordenadores interpretarlo. En el caso del estándar JATS se utiliza el metalenguaje XML. XML a su vez tomó prestado lo que hasta el momento utilizaba SGML para ese propósito: la definición de tipo de documento, *Document Type Definition* (DTD), el cual es un lenguaje con un vocabulario y una sintaxis específica, diseñado para definir con rigor los modelos de documento.

Surgen posteriormente también otros lenguajes de esquema que son útiles para este mismo propósito, como *XML Schema Language*, publicado como recomendación de W3C en 2001. Este lenguaje también se utiliza para describir de una forma precisa la estructura y las restricciones en los contenidos de los documentos XML, y posee además características que no están disponibles en las DTD. RELAX NG es otro de los lenguajes de esquema que puede utilizarse para definir modelos, pues hace básicamente lo mismo que los dos lenguajes mencionados anteriormente, pero con una sintaxis diferente, más simple que la de *XML Schema*. Fue desarrollado por ISO/IEC y su primera versión fue publicada en 2003 como estándar internacional ISO/IEC 19757-2.

El diseño de un esquema se basa un modelo, el cual es el resultado del análisis de grupos o clases de documentos que comparten determinadas características. Kasdorf (2003) plantea lo siguiente: "*Document analysis is the foundation on which a formal Document Type Definition (DTD) or schema is built. The most basic factor that affects the complexity of the document analysis process is*

---

<sup>1</sup> Se ha preferido utilizar los términos *Tag Suite* y *Tag Sets* en inglés, tal cual aparecen en el documento oficial del estándar NISO JATS, para referirnos al conjunto completo de elementos y atributos XML predefinidos, y a los conjuntos de etiquetas, respectivamente.



*the scope of the desired definition. Although it is possible-and can be powerful- to write a definition specific to a single tile, most DTD or schemas are created for classes of documents: groups of documents that share certain characteristics. Usually, that means they share a common structure."*

Desarrollar un esquema que acomode diferentes tipos de documentos en una sola definición resulta en un gran conjunto de elementos, con etiquetas muy neutras y genéricas que terminan siendo complejas de manejar para editoriales, autores y diseñadores. Las llamadas DTD o esquemas "de archivo" pueden abarcar toda la variedad de tipos de documentos de una editorial. Sirven al doble propósito de organizar todo el acervo de la editorial, y se pueden utilizar además para volcar todo o parte de ese acervo a un sistema de gestión de contenidos (CMS). Lo que sucede generalmente es que esas definiciones basadas en un alcance excesivamente amplio resultan ser complejas y abstractas, por lo cual se utilizan conjuntamente con varias definiciones de propósito específico (las llamadas DTD de base) que están diseñadas justamente para tipos específicos de publicaciones. La solución que plantea Kasdorf (2003) a esta problemática es desarrollar una *Suite* de definiciones, o bien una definición con componentes modulares.

Todos los estándares, incluido JATS, adoptan DTD externas cuando utilizan DTD como lenguaje de esquema para expresar los modelos de documento especificados en el estándar. Las DTD externas se declaran en un archivo externo .dtd y la declaración de tipo de documento del fichero XML `<!DOCTYPE>` debe contener la referencia al archivo DTD, tal como se indica en el siguiente ejemplo (Figura 2).

```
<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal
Publishing DTD v1.0 20120330//EN" "JATS-
journalpublishing1.dtd">
```

Figura 2. La referencia al archivo .dtd en un documento XML marcado con JATS

Fuente: Adaptado de Huh (2014)

Un documento XML está estructurado en dos grandes bloques: prólogo e instancia. Los constructores que interesan desde el punto de vista de la creación de una DTD son los que forman parte de la instancia del documento: elementos, atributos, entidades, PCDATA, CDATA. Los elementos son los principales bloques de construcción y los atributos proporcionan información adicional sobre los elementos. Las entidades pueden ser consideradas como abreviaciones para otros contenidos, de forma que para las entidades definidas en una DTD, el contenido es sustituido cuando el software encuentra la entidad mencionada en el fichero XML. El uso de entidades en las DTD, especialmente el uso de **entidades de parámetro**, simplifica la escritura de DTD complejas que usan a menudo las mismas estructuras varias veces. PCDATA y CDATA son tipos de datos. PCDATA significa *Parsed Character Data* y es texto que será analizado por un *parser* (analizador sintáctico), que analiza las entidades y las marcas. De esta forma el proceso de validación asegura que las etiquetas en el interior del texto se tratarán como indica el marcado y las entidades se expandirán/sustituirán de acuerdo a lo definido en el esquema. CDATA significa *Character Data* y se trata de texto que no será analizado por el *parser*.

*JATS DTD Suite*<sup>2</sup> ha sido escrita como una serie de módulos XML que se pueden combinar para ser utilizados en los diferentes conjuntos de etiquetas (*Tag Sets*). Los módulos son archivos físicamente separados, los cuales en diferentes combinaciones definen todas las estructuras en que se organizan los elementos (como tablas, fórmulas químicas y matemáticas, párrafos, secciones, figuras, notas al pie y referencias), atributos y entidades de la *Suite*.

*An XML DTD module consists of a set of element types, a set of attribute list declarations, and a set of content model declarations, where any of these three sets may be empty. An attribute list declaration in an XML DTD module may modify an element type outside the element types in the module, and a content model declaration may modify an element type outside the element type set (W3C, 2000).*

En las DTD modulares, la DTD está compuesta de muchos módulos diferentes (normalmente ficheros. ent o .mod) de manera que las piezas de la DTD se pueden reutilizar, ignorar o incluir fácilmente. Los módulos son normalmente referenciados (invocados) desde una DTD de base, utilizando entidades de parámetro (*parameter entities*) o bien referencias a otros archivos que contienen declaraciones de elementos, de listas de atributos, de entidades, etc. En el caso de JATS se utilizan entidades de parámetro para construir DTD más flexibles, permitiendo definir un modelo básico y luego reutilizar ese modelo varias veces, e incluso construir otros modelos reutilizando partes del modelo básico.

Una entidad de parámetro es una construcción sintáctica XML que permite nombrar una colección de elementos, atributos o valores de atributos, para que puedan ser referidos por su nombre en lugar de tener que listar todos sus componentes cada vez que necesiten ser utilizados en una sentencia XML. Los nombres de las entidades de parámetro siempre comienzan con un signo de porcentaje y terminan con punto y coma. Ejemplo: la denominación **%abbrev-journal-title-elements**; corresponde a la entidad **Abbreviated Journal Title Elements** definida en el módulo **JATS-journalmeta1.ent** y contiene el siguiente fragmento de un modelo de contenido reutilizable para diversos elementos (Figura 3):

```
| email | ext-link | uri | inline-supplementary-material |
related-article | related-object | hr | bold | italic | monospace
| overline | overline-start | overline-end | roman | sans-serif
| sc | strike | underline | underline-start | underline-end |
alternatives | inline-graphic | private-char | chem-struct |
inline-formula | tex-math | mml:math | abbrev | milestone-end |
milestone-start | named-content | styled-content | fn | target
| xref | sub | sup | x
```

Figura 3. Modelo de contenido de la entidad *Abbreviated Journal Title Elements*  
Fuente: módulo JATS-journalmeta1.ent

Las entidades de parámetro utilizadas en *JATS DTD Suite* pueden ser: de clase, de mezclas, de modelos de contenido o listas de atributos. Las entidades de clase, constituidas por grupos de elementos separados por “|”(barra OR) llevan en su nombre el sufijo “.class”. Ejemplo: `<!ENTITY % list.class "def-list | list">`. Las entidades de tipo mezcla (grupos de clases separados por |) no tienen sufijos, solo algunas han sido nombradas “.mix”. Algunas entidades de este tipo se utilizan

<sup>2</sup> *JATS DTD Suite* es una expresión frecuente en la documentación técnica de JATS, utilizada para referirse al conjunto de módulos combinables en que se ha fraccionado la DTD con el fin de servir a distintos propósitos.

conjuntamente con #PCDATA en un modelo de contenido mixto y todas comienzan con "|". Ejemplo: `<!ENTITY % rendition-plus "| %emphasis.class; | %subsup.class;"` >. Las entidades de parámetro que describen modelos de contenido o sobrescritura de modelos de contenido, usan contenido mixto y clases para todos los grupos OR (grupos separados por la barra OR). Las entidades de sobrescritura de modelos de contenido son de dos tipos: modelos y elementos. Se describen de forma separada para preservar el contenido mixto, o la naturaleza del contenido de los elementos que se convierten de otros modelos, como ayuda para el intercambio. Las del tipo modelos se encargan de sobrescribir un modelo de contenido completo, y son nombradas con el sufijo "-model", como lo muestra el siguiente ejemplo (Figura 4):

```
<!ENTITY % kwd-group-model
    "(label?, title?,
    ((%kwd.class; | %x.class;)+ |
    unstructured-kwd-group ) )">
```

Figura 4. Ejemplo de entidad de parámetro para sobrescritura de modelos  
Fuente: módulo JATS-archivecustom-models1.ent

Las entidades del tipo elementos sobrescriben agrupaciones de elementos que pueden además estar mezclados con #PCDATA dentro de un modelo de contenido mixto. Son nombradas con el sufijo "-elements". Ejemplo: `"%copyright-holder-elements;"`. Todas las entidades que sobrescriben agrupaciones de elementos comienzan con una barra de OR, de modo que un modelo pueda excluir todos los elementos y ser reducido a #PCDATA de ser necesario, como se puede apreciar en el siguiente ejemplo (Figura 5):

```
<!ENTITY % copyright-holder-elements
    "| %institution-wrap.class; | %subsup.class; |
    %x.class;">
```

Figura 5. Ejemplo de entidad de parámetro para sobrescritura de grupos de elementos  
Fuente: módulo JATS-archivecustom-models1.ent

Las entidades de parámetro que definen listas de atributos para un elemento particular son del tipo listas de atributos. Son nombradas con el nombre del elemento, seguido por el sufijo "-atts". En el siguiente ejemplo se aprecian los atributos para el elemento `<abstract>`, los cuales tienen por nombre "abstract-atts" (Figura 6).

```
<!ENTITY % abstract-atts
    "%jats-common-atts;
    abstract-type CDATA          #IMPLIED
    specific-use CDATA           #IMPLIED
    xml:lang NMTOKEN             #IMPLIED">
```

Figura 6. Ejemplo de entidad de parámetro que describe una lista de atributos  
Fuente: módulo JATS-articlemeta1.ent

## 4.2. La Suite de módulos JATS

Según el sitio de NLM dedicado al estándar, "*JATS is an application of NISO Z39.96-2012, which defines a set of XML elements and attributes for tagging journal articles and describes three article models.*" Como ya se ha comentado, el conjunto de elementos y atributos XML es lo que aparece referido como *Tag Suite* o simplemente *Suite* en el texto del estándar, y los tres modelos de artículo que forman parte del estándar son los tres conjuntos de etiquetas o *Tag Sets*.

Cada elemento de la *Suite* es identificado con una etiqueta. La etiqueta o "*tag*" es el nombre corto del elemento, el cual es utilizado en los documentos XML, y en DTD y esquemas. Los nombres de las etiquetas están rodeados por paréntesis angulares, como establece el metalenguaje XML. Por ejemplo, en JATS DTD el nombre de etiqueta <p> se utiliza para el elemento *Paragraph*. Muchos elementos de la *Suite* son utilizados en más de un *Tag Set*, otros son utilizados en todos los *Tag Sets* y otros elementos no son utilizados en ninguno de los 3 *Tag Sets*, pero se incluyen en la *Suite* como "*building blocks*" para la creación de nuevos *Tag Sets* o posibles ampliaciones de los existentes.

Cada atributo se identifica con un nombre XML (nombre que se utiliza en el marcado de los documentos XML). La *Suite* define los tipos y las listas de valores posibles o recomendados de los atributos. Algunos atributos se utilizan en muchos elementos. Otros atributos se utilizan de la misma manera en todos los elementos de un conjunto de etiquetas, pero algunos pueden tener diferente opcionalidad o diferentes contenidos en distintos elementos del mismo conjunto de etiquetas.

En la DTD, cada elemento de la *Suite* es definido en una declaración de elemento, que especifica qué contiene o debe contener el elemento, en qué orden, y cuántos. A su vez, cada atributo es definido en las declaraciones de atributo. Allí se indica qué atributos puede o debe tener un elemento. También las entidades deben ser definidas antes de poder ser utilizadas.

*JATS DTD Suite* ha sido escrita como un conjunto de módulos XML, cada uno de los cuales es un fichero independiente. Ningún módulo es un esquema entero por sí mismo, sino que esos módulos se pueden combinar para dar lugar a varios esquemas diferentes. Es así como a partir de la *Suite* se construyen los tres esquemas o *Tag Sets* que incluye el estándar NISO JATS. Además, la *Suite* está pensada para la creación de nuevos esquemas a partir de ella<sup>3</sup>.

El diseño modular de la *Suite* se basa en diferentes tipos de módulos, dependiendo de su función. Los módulos comunes son utilizados por todos los *Tag Sets*. Algunos módulos sirven para definir los elementos dentro de una estructura particular, nombrando todos los potenciales componentes de una lista. Hay módulos que identifican a los miembros de una "clase" de elementos, donde la clase es una agrupación de elementos que comparten un uso similar. Básicamente, *JATS DTD Suite* utiliza cinco tipos de módulos: módulos comunes, módulos que definen elementos y atributos que componen la *Suite*, módulos que definen elementos y atributos MathML 2.0 y MathML 3.0, módulos que definen el modelo tabla inspirado en XHTML y módulos que definen entidades de caracteres y notaciones. Los dos primeros tipos de módulos pueden ser sobrescritos por los módulos personalizados de cada *Tag Set*. Los tres últimos solo necesitan ser invocados si el esquema los utiliza. En las Tablas 2 a la 6 se presentan agrupados los 5 tipos de módulos, con el nombre de fichero de cada módulo y su respectiva función dentro de la Suite.

---

<sup>3</sup> How to Build A New Custom DTD <http://dtd.nlm.nih.gov/#custom>

El módulo de módulos (*JATS DTD Suite Module of Modules*) es pieza clave del diseño de la *Suite*, pues contiene las declaraciones de entidades y de todos los módulos externos de *JATS DTD Suite*. Los módulos no están referenciados, o sea no son invocados en este módulo de módulos, sino que simplemente están declarados (nombrados). Son las DTD de base de cada *Tag Set*, o un módulo de configuración como por ejemplo el de las tablas XHTML, los que invocan cada módulo. Con esta lógica de diseño, cuando se quiere incluir un conjunto de elementos en un esquema determinado, basta hacer referencia a la entidad externa de parámetro del módulo que contiene las declaraciones del conjunto deseado para modificar las clases o modelos de contenido, o bien para utilizar nuevos elementos. Es importante mencionar que en este diseño existe un orden para invocar los diferentes módulos de la *Suite*, ya que las entidades de parámetro necesitan ser definidas antes de ser utilizadas.

Además del módulo de módulos, existen otros tres módulos que conjuntamente con el anterior son del tipo módulos comunes de *JATS DTD Suite* (Tabla 2): el módulo de clases de elementos por defecto, el módulo de mezclas de elementos por defecto y el módulo de elementos comunes (compartidos). El módulo de clases de elementos por defecto (*Default Element Classes Module*) declara las entidades de parámetro que se utilizan para definir las clases de elementos por defecto, o sea, define las clases de elementos a ser utilizadas en los demás módulos de *JATS DTD Suite*. El módulo de mezclas de elementos por defecto (*Default Element Mixes Module*) declara los valores por defecto de todas las mezclas de elementos utilizados en los modelos de contenido de la DTD. El módulo de elementos comunes compartidos (*Common Elements Module*) define las entidades de parámetro comunes, invoca los módulos compartidos (como el módulo de caracteres especiales, *Standard XML Special Characters Module* o el módulo de notaciones, *Notation Declarations*) y declara elementos que no encajan en una clase en particular pues serán utilizados en más de un nivel estructural.

Tabla 2. Módulos comunes de *JATS DTD Suite*  
Fuente: elaboración propia.

Nombre del módulo	Fichero de módulo	Función del módulo
<i>Suite Module of modules</i>	JATS-modules1.ent	Nombra todos los módulos externos (excepto a sí mismo o a los módulos de personalización) que son parte de la <i>JATS DTD Suite</i> .
<i>Suite Default Element Classes Module</i>	JATS-default-classes1.ent	Declara las entidades de parámetro utilizadas para definir las clases de elementos por defecto.
<i>Suite Default Element Mixes Module</i>	JATS-default-mixes1.ent	Declara los valores por defecto para todas las mezclas de elementos utilizados en los modelos de contenido de la <i>Suite</i> .
<i>Common (Shared) Elements Module</i>	JATS-common1.ent	Define las entidades de parámetro comunes, invoca los módulos compartidos (como el de caracteres especiales y notaciones) y declara los elementos que no encajan en ninguna clase.

Tabla 3. Módulos que definen elementos y atributos en JATS DTD Suite

Fuente: elaboración propia.

Nombre del módulo	Fichero de módulo	Función
<i>Journal Article Metadata Elements Module</i>	JATS-articlemeta1.ent	Declara los elementos de metadatos utilizados para describir un artículo de revista, incluyendo tanto metadatos del número como elementos de la cabecera del artículo. Nombra todos los elementos que se utilizan para describir la revista en la que se ha publicado el artículo.
<i>Back Matter Elements Module</i>	JATS-backmatter1.ent	Declara los elementos que no forman parte del texto principal del artículo, pero se consideran materiales auxiliares tales como apéndices, glosarios y listas de referencias bibliográficas.
<i>Display Class Elements Module</i>	JATS-display1.ent	Declara los elementos como figuras, gráficos, expresiones matemáticas, estructuras y formulas químicas, tablas, etc. para imprimirlos o mostrarlos en pantalla.
<i>Format Class Elements Module</i>	JATS-format1.ent	Declara elementos relacionados con los elementos de <i>Appearance Class</i> , <i>Break Class</i> y <i>Emphasis Class</i> .
<i>Funding Elements Module</i>	JATS-funding1.ent	Declara los elementos relacionados con acceso abierto, auspiciantes, y otra información sobre el origen de los fondos que hicieron posible la investigación.
<i>Journal Metadata Elements Module</i>	JATS-journalmeta1.ent	Declara los elementos utilizados para describir la revista en la que se ha publicado el artículo.
<i>Link Class Elements Module</i>	JATS-link1.ent	Declara elementos que son enlaces, internos o externos, como las URL (<uri>) o las referencias internas (<xref>).
<i>List Class Elements Module</i>	JATS-list1.ent	Declara los elementos de la lista de clases ( <i>List Class</i> ). Son declaradas aquí todas las listas excepto la de referencias bibliográficas.
<i>Math Class Elements Module</i>	JATS-math1.ent	Declara los elementos de las clases relacionadas con expresiones matemáticas, como las ecuaciones.
<i>NLM Citation Module</i>	JATS-nlmcitation1.ent	Agrega el modelo NLM para estructurar las referencias bibliográficas (<nlm-citation>), pero es un elemento obsoleto, que debe ser remplazado por <element-citation> o por <mixed-citation>.
<i>Paragraph-Like Elements Module</i>	JATS-para1.ent	Declara elementos estructurales que pueden aparecer en los párrafos; estos elementos son nombrados en las diversas entidades de parámetro relacionadas con los párrafos.
<i>Subject Phrase Class Elements Module</i>	JATS-phrase1.ent	Declara los elementos de la clase <i>Phrase Class</i> , que nombra los elementos sobre temas específicos.
<i>Bibliographic Reference (Citation) Class Elements Module</i>	JATS-references1.ent	Declara los elementos de las referencias bibliográficas.
<i>Related Object Elements Module</i>	JATS-related-object1.ent	Define el elemento contenedor <related-object>, el cual se utiliza como contenedor para enlaces de texto y descripciones de objetos externos como por ejemplo una hoja de cálculo o un libro relacionado con el artículo.



<i>Section Class Elements Module</i>	JATS-section1.ent	Declara todos los elementos a nivel de sección en <i>JATS DTD Suite</i> .
<i>OASIS XML Exchange (CALS) Table Setup Module</i>	JATS-oasis-tablesetup1.ent	Establece todas las entidades de parámetro que necesita el modelo de tabla <i>OASIS XML Exchange (CALS)</i> e invoca el módulo que contiene dicho modelo.
<i>OASIS XML Exchange (CALS) Table Model Module</i>	oasis-exchange.ent	Es el módulo <i>OASIS XML Exchange (CALS) table model</i> . Este módulo es invocado en %oasis-tablesetup.ent;.

Tabla 4. Módulos que definen elementos y atributos MathML 2.0 y MathML 3.0

Fuente: elaboración propia.

Nombre del módulo	Fichero de módulo	Función
<i>MathML 2.0 Setup Module</i>	JATS-mathmlsetup1	Este módulo configura todas las entidades de parámetro necesarias para utilizar el <i>Tag Set MathML 2.0</i> e invoca el módulo <i>MathML 2.0 Tag Set</i> que a su vez invoca todos los demás módulos <i>MathML 2.0</i> .
<i>MathML 3.0 Setup Module</i>		Las mismas funciones que <i>MathML 2.0</i> , por lo cual no pueden usarse ambos a la vez en un mismo <i>Tag Set</i> .
<i>MathML DTD</i>	mathml2.dtd	Se utilizan para el Lenguaje de Marcado Matemático (MathML) 2.0 o MathML 3.0 respectivamente. Su función es describir las notaciones matemáticas, capturando su estructura y también su contenido.
<i>MathML Qualified Names 1.0</i>	mathml2-qname-1.mod	Declara las entidades de parámetro para los espacios de nombre, el prefijo para MathML y las entidades de parámetro utilizadas para proporcionar los espacios de nombre a todos los elementos MathML.
<i>Extra Entities for MathML</i>	mmlextra.ent	Se utiliza para el procesamiento de MathML, tanto para MathML 2.0 como para MathML 3.0.
<i>Aliases for MathML</i>	mmlalias.ent	Se utiliza para el procesamiento de MathML, tanto para MathML 2.0 como para MathML 3.0.

Tabla 5. Módulos que definen el modelo tabla inspirado en XHTML

Fuente: elaboración propia.

Nombre del módulo	Fichero de módulo	Función
<i>NISO JATS XHTML-inspired Table Setup Module</i>	JATS-XHTMLtablesetup1	Establece todas las entidades de parámetro que necesita el módulo inspirado en el modelo de tabla XHTML NISO JATS, y luego invoca el módulo que contiene ese modelo. Para incluir el modelo de tabla XHTML NISO JATS en un conjunto de etiquetas, su DTD debe hacer referencia a este módulo.

NISO JATS XHTML-inspired Table Module	xhtml-table-1.mod	Es la versión pública del módulo inspirado en el modelo tabla XHTML. Este es el modelo de tabla por defecto para la <i>Suite</i> .
NISO JATS XHTML-inspired Table Style Module	xhtml-inlstyle-1.mod	Declara los atributos @style que permiten el estilo de marcado para elementos como <td> y <tr> dentro del modelo tabla XHTML.

Tabla 6. Módulos que definen entidades de caracteres y notaciones (todos opcionales)  
Fuente: elaboración propia.

Nombre del módulo	Fichero de módulo	Función
XML Special Characters Module	JATS-xmlspecchars1.ent	Estándar ISO XML para entidades de caracteres especiales.
Custom Special Characters Module	JATS-chars1.ent	Entidades de caracteres especiales creados específicamente para su uso en <i>JATS DTD Suite</i> .
Notation Declarations Module	JATS-notat1.ent	Módulo contenedor para las <i>Notation Declarations</i> que se utilizan en esta <i>Suite</i> .

Las ventajas que se mencionan en el sitio web NISO JATS respecto al enfoque modular de diseño de la *Suite* son varias. La principal ventaja es que las unidades estructurales más pequeñas se escriben una vez, haciendo más fácil mantener consistencia en las estructuras de bajo nivel para todos los tipos de documento. Además, un *Tag Set* para una nueva función, o un nuevo tipo de publicación, se puede construir de forma rápida, ya que la mayoría de los componentes necesarios ya están definidos en la *Suite*. Otra ventaja es que el software que se haya adaptado para un proyecto (ya sea orientado a creación, edición, visualización de documentos, etc.) se puede compartir en otros proyectos, modificando solo las partes que verdaderamente lo requieran. Además, el personal de edición y producción puede capitalizar la experiencia adquirida en un proyecto de marcado y aplicarla directamente a otro proyecto sin necesidad de una nueva capacitación.

Según Kassdorf (2003), una ventaja del enfoque modular de diseño es que las definiciones se pueden ir desarrollando a lo largo del tiempo. De esa forma se puede crear primero una DTD para optimizar la publicación en web y la impresión de artículos. Esa DTD puede estar compuesta por otros módulos subordinados que permitan describir otros tipos de documentos que se publican en una revista. Todo esto sin perjuicio que se puedan desarrollar más adelante otras DTD para monografías, atlas u otros tipos de documentos. Luego se puede desarrollar incluso una DTD “maestra” que contenga todas las clases de documentos.

La mayor desventaja que se menciona acerca del sistema modular en el sitio web NISO JATS es la existencia de una curva de aprendizaje más larga cuando se capacita al personal, pues no resulta obvio dónde encontrar un elemento o atributo en particular utilizando un sistema modular.



### 4.3. Modelos específicos: *Tag Sets* o conjuntos de etiquetas

Los conjuntos de etiquetas o *Tag Sets* son modelos descriptivos para las estructuras XML de los artículos, basados en los componentes que define la *Suite*. El estándar JATS incluye tres modelos o conjuntos, los cuales ya han sido introducidos brevemente en el capítulo 3: *Journal Archive and Interchange Tag Set*, *Journal Publishing Tag Set* y *Article Authoring Tag Set*. Están diseñados para funciones específicas: archivo e intercambio, publicación y autoría, respectivamente.

Los *Tag Sets*, además de utilizar los módulos de la *Suite*, poseen cada uno módulos específicos: un módulo principal o módulo de base, y varios módulos subordinados. De esta manera se habilita la sobrescritura de las instrucciones de los módulos comunes de la *Suite*, aplicando así personalizaciones que los hacen adecuados a los usos concretos para los cuales han sido destinados.

En las denominadas "*Tag Libraries*", la descripción formalizada en cada *Tag Set* (mediante DTD, XSD o RNG), se expresa además en lenguaje natural e incluye la descripción de todos los elementos que contiene cada modelo, identificados por la etiqueta y el nombre correspondiente (Ej.: etiqueta `<trans-title>`, nombre de la etiqueta *Translated Title*). Cada *Tag Set* especifica las relaciones entre elementos o grupos de elementos, por ejemplo: "*Any combination of*" (si las estructuras pueden aparecer en cualquier secuencia), o "*The following, in order*" (si las estructuras deben aparecer en un orden determinado). También en cada *Tag Set* aparecen especificados los atributos que un elemento puede tener, incluyendo el nombre del atributo y el nombre del atributo escrito en forma extendida. Los nombres de atributos están precedidos por el símbolo @ (Ej.: nombre del atributo `@abstract-type`, nombre extendido del atributo: *Type of Abstract*). Para algunos atributos se proporciona información adicional sobre cómo debe ser utilizado o interpretado. Los atributos pueden ser utilizados por un sólo elemento, o puede haber más de un elemento asociado a un atributo. Hay casos en que el uso de un atributo es el mismo en todos los elementos del conjunto. Además, la opcionalidad o no de un elemento o de un atributo puede variar en los diferentes *Tag Sets*. En las respectivas *Tag Libraries* se presentan todos los elementos asociados con un atributo para cada *Tag Set*, se indica si es opcional o no, y una tabla muestra el valor o los valores permitidos del atributo y describe el significado de cada valor. Si hay un valor por defecto, ese valor aparece en negrita en la tabla.

En el caso de JATS, cuando se descarga un *Tag Set* desde el sitio FTP de NLM, lo que se obtiene al descomprimir el fichero ZIP son todos los componentes del *Tag Set* en un solo directorio. En otras palabras, una vez descomprimido el fichero ZIP, se creará en nuestro ordenador un directorio conteniendo todos los módulos, tanto los comunes de la *Suite* como los específicos del *Tag Set* que se ha descargado. Para validar un artículo en local (utilizando aplicaciones de software instaladas en el ordenador, como XML Spy) se debe indicar al *parser* cuál es la DTD de base del *Tag Set*.

Como uno de sus propios desarrolladores lo reconoce, "*An intimidating number of variations are available for JATS version 1.1*" (Usdin, 2016). Esto se debe a que en la última versión del estándar JATS, los tres *Tag Sets* tienen "variaciones", las cuales son idénticas excepto en lo que respecta al modelado de las tablas y al modelado de las fórmulas matemáticas. A su vez, cada una de esas variaciones se continúan entregando en las tres sintaxis en las que ya se proporcionaba la versión 1.0: DTD, XSD (W3C XML Schema), and RNG (RELAX NG). Esto implica que cada usuario de esta versión debe escoger una variante según las tablas o el lenguaje para fórmulas que mejor se adapte a sus necesidades, y proceder entonces a descargar la variante de la versión correspondiente a la sintaxis que haya elegido utilizar.

#### 4.3.1. *Journal Archive and Interchange Tag Set* / Conjunto de etiquetas de archivo e intercambio

*Journal Archiving and Interchange Tag Set* (JATS-archivearticle1.dtd y sus módulos específicos de personalización) definen un *Tag Set* centrado en la función de archivo (DTD para repositorios) e intercambio, si bien el conjunto de módulos también puede servir de base para la publicación.

La característica principal de este *Tag Set* es que posibilita a un archivo capturar componentes semánticos y estructurales de materiales existentes, sin modelar ninguna secuencia ni formato textual en particular. Uno de los principales objetivos de este conjunto de etiquetas es preservar el contenido intelectual de los artículos de revistas, independientemente de la forma en que el contenido haya sido generado originalmente. Es el más permisivo de los conjuntos de etiquetas, pues pretende ser muchas cosas al mismo tiempo: *Tag Set* de base para repositorios XML, destino para la conversión desde otros conjuntos de etiquetas, y formato común para el intercambio basado en XML entre editoriales, archivos, agregadores y proveedores de servicios.

El conjunto de etiquetas de archivo e intercambio está diseñado para la conversión desde múltiples fuentes, con finalidad de unificar otras estructuras de marcado XML en un único modelo. Ese enfoque ha hecho de este *Tag Set* un conjunto extenso e inclusivo. Muchos elementos han sido creados expresamente para que la información etiquetada no se pierda cuando se convierte material desde otro conjunto de etiquetas a este *Tag Set*. Su diseño también incluye varios mecanismos (con frecuencia utilizando atributos) para preservar el contenido intelectual de la estructura de un documento cuando esa estructura se convierte a JATS, incluso cuando no hubiese en JATS ningún elemento equivalente a la estructura de origen.

Los módulos específicos del conjunto de archivo e intercambio son cinco, agrupados en tres tipos de módulos: módulo de base, módulo de módulos y tres módulos de personalización (Tabla 7). Estos módulos se encargan de los cambios en las declaraciones por defecto de la *Suite*. Las entidades son redefinidas en los módulos personalizados, que son invocados desde el módulo principal de *Journal Archiving and Interchange DTD*: JATS-archivearticle1.dtd.

Tabla 7. Módulos específicos de *Journal Archiving and Interchange DTD* (v1.0 20120330)  
Fuente: elaboración propia.

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	JATS-archivearticle1.dtd	<i>Journal Archiving and Interchange DTD module</i>	Módulo principal, o módulo de base del <i>Tag Set</i> . Todos los demás módulos específicos están subordinados al módulo de base. Su propósito es declarar las entidades de parámetro que se usan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>JATS DTD Suite</i> .
<i>Module to DTD-specific modules</i>	JATS-archivecustom-modules1.ent	<i>Journal Archiving and Interchange DTD module of modules</i>	Este módulo nombra todos los módulos creados expresamente para esta DTD, los cuales no están presentes en <i>JATS DTD Suite</i> . Declara las entidades de parámetros externas para todos los módulos específicos utilizados por la DTD de archivo e intercambio. Los módulos solo se declaran en este módulo, y se invocan desde el módulo principal del <i>Tag Set</i> .

Customization modules	JATS- archivecustom- classes1.ent	<i>Journal Archiving DTD Customize Classes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de JATS DTD Suite para las clases de elementos. Como las entidades de parámetro deben ser declaradas antes de ser utilizadas, este módulo es invocado en la DTD de base antes que los módulos de contenido que declaran los elementos, y antes del módulo de clases predeterminadas.
	JATS- archivecustom- mixes1.ent	<i>Journal Archiving DTD Customize Mixes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de JATS DTD Suite para las mezclas de elementos de propósito general. Este módulo se invoca antes que <i>Suite Default Element Mixes Module</i> en la DTD de base.
	JATS- archivecustom- models1.ent	<i>Journal Archiving DTD Customize Content and Attributes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de JATS DTD Suite. En este módulo, los grupos OR dentro de los modelos utilizan mezclas o clases en lugar de utilizar elementos directamente.

En la versión actualmente vigente del estándar JATS (v. 1.1), hay cuatro variantes de la DTD de Archivo e Intercambio: Archivo e Intercambio que utiliza tablas XHTML con MathML 2.0; Archivo e Intercambio que utiliza tablas XHTML con MathML 3.0; Archivo e Intercambio que utiliza tablas XHTML y tablas OASIS CALS con MathML 2.0; Archivo e Intercambio que utiliza tablas XHTML y tablas OASIS CALS con MathML 3.0. Cada versión de la DTD se corresponde con un fichero ZIP en el FTP, existiendo cuatro ficheros separados que contienen las diferentes variantes de la DTD de Archivo e Intercambio, debiendo el usuario elegir una de esas DTD en base a sus requisitos para tablas y MathML.

La última versión oficial de JATS *Archiving and Interchange Tag Set* (versión 1.1, 2015) está disponible en Internet para su descarga en: <ftp://ftp.ncbi.nlm.nih.gov/pub/jats/archiving/1.1/>

#### 4.3.2. Journal Publishing Tag Set / Conjunto de etiquetas de publicación

*Journal Publishing Tag Set* define los elementos y atributos que describen el contenido y los metadatos de artículos de revistas, incluyendo artículos de investigación y otros tipos de artículos, además de cartas, editoriales, reseñas de libros y comentarios de productos. La DTD de Publicación ha sido desarrollada en base a la DTD de Archivo e Intercambio, pero optimizada para servir de formato común para la generación y publicación de contenidos de revista en XML. Al igual que el *Tag Set* de archivo e intercambio, el *Tag Set* de publicación también permite la descripción del contenido del artículo completo o sólo de los metadatos de cabecera del artículo.

Es un conjunto de etiquetas moderadamente prescriptivo, diseñado para los archivos que necesitan normalizar y controlar su contenido. Este *Tag Set* también está pensado para ser utilizado por los editores en la implementación del marcado XML de contenidos de revista que han sido generalmente convertidos desde otros formatos como por ejemplo PDF o *Microsoft Word*.

La filosofía de *Publishing Tag Set* es preferir una sola forma estructural siempre que sea posible. Los elementos y las opciones de marcado han sido limitados a los efectos de obtener estructuras de datos que faciliten la generación de productos y proporcionen una única ubicación de la información para la búsqueda. Es así que este *Tag Set* es más pequeño que el *Tag Set* de archivo e intercambio pues contiene menos elementos, menos opciones en muchos contextos, y se imponen más a menudo secuencias particulares a los elementos. El establecimiento de secuencias de elementos tiene como finalidad facilitar la publicación en web y también la impresión de los artículos.

La DTD de publicación se ha desarrollado a partir de los módulos de *JATS DTD Suite*, realizando cambios en las declaraciones de módulos de la *Suite* mediante la sobrescritura de los contenidos de sus entidades de parámetro. Las entidades son redefinidas en los módulos personalizados de publicación, que son invocados desde el módulo principal de *Publishing DTD*: JATS-journalpublishing1.dtd. Todos los módulos anteriormente mencionados son los que definen este *Tag Set*, cuyos nombres y funciones se presentan en la Tabla 8.

Tabla 8. Módulos específicos de *Journal Publishing DTD*  
Fuente: elaboración propia

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	JATS-journalpublishing1.dtd	<i>Journal publishing DTD module</i>	Módulo principal, o módulo de base del <i>Tag Set</i> : en él se pueden habilitar y deshabilitar los demás módulos específicos, todos subordinados al módulo de base. Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>JATS DTD Suite</i> .
<i>Module to name DTD-specific modules</i>	JATS-journalpubcustom-modules1.ent	<i>Journal publishing DTD module of modules</i>	Este módulo nombra todos los módulos creados expresamente para esta DTD, los cuales no están presentes en <i>JATS DTD Suite</i> . Declara las entidades de parámetro externas para todos los módulos específicos utilizados por la DTD de publicación. Los módulos solo se declaran en este módulo; se invocan desde el modulo principal del <i>Tag Set</i> .
<i>Customization modules</i>	JATS-journalpubcustom-classes1.ent	<i>DTD-specific class customizations module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>JATS DTD Suite</i> para las clases de elementos.
	JATS-journalpubcustom-mixes1.ent	<i>DTD-specific mix customizations module</i>	Configura las entidades de parámetro y las definiciones de clases elementos que se utilizan para sobrescribir ciertas mezclas de elementos de <i>JATS DTD Suite</i> .
	JATS-journalpubcustom-models1.ent	<i>DTD-specific models/attributes customizations module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>JATS DTD Suite</i> .

En la versión actualmente vigente del estándar JATS (v. 1.1), hay cuatro variantes de la DTD de Publicación: Publicación que utiliza tablas XHTML con MathML 2.0; Publicación que utiliza tablas XHTML con MathML 3.0; Publicación que utiliza tablas XHTML y tablas OASIS CALS con MathML 2.0; Publicación que utiliza tablas XHTML y tablas OASIS CALS con MathML 3.0. Cada versión de la DTD se corresponde con un fichero ZIP en el FTP, existiendo cuatro ficheros separados que contienen las diferentes variantes de la DTD de Publicación, debiendo el usuario elegir una de esas DTD en base a sus requisitos para tablas y MathML.

La última versión oficial de JATS *Article Publishing Tag Set* (versión 1.1, 2015) está disponible en Internet para su descarga en: <ftp://ftp.ncbi.nlm.nih.gov/pub/jats/publishing/1.1/>

#### 4.3.3. *Article Authoring Tag Set* / Conjunto de etiquetas de autoría

*Article Authoring Tag Set* es un conjunto de etiquetas optimizado para la autoría de nuevos artículos de revista, el cual puede ser utilizado por los autores para presentar sus contribuciones a revistas o para depositarlos en repositorios como *PubMed Central*. Aunque en teoría el alcance del *Tag Set* es el mismo que para el conjunto de etiquetas de publicación, en la práctica *Article Authoring Tag Set* define elementos y atributos centrados en artículos de investigación.

Mientras que la DTD de *Archivo e Intercambio* puede tener muchas formas de expresar la misma información, *Article Authoring Tag Set* es más prescriptivo que descriptivo e incluye muchos elementos cuyos contenidos deben ocurrir en un orden específico. Es un conjunto restringido, orientado a la creación de contenido y cuya intención no es limitar el poder expresivo de la DTD, sino limitar las opciones de marcado. En una DTD de archivo e intercambio tener muchas opciones de marcado tiene mucho sentido, pues permite la conversión desde una amplia variedad de formatos, pero para la autoría de artículos no son necesarias tantas opciones. Para este *Tag Set*, la prioridad es lograr una normalización y un control mucho mayor de los contenidos, por lo tanto, la filosofía de este *Tag Set* es definir una sola manera de marcar la estructura del artículo.

Otras cuestiones que tienen que ver específicamente con la autoría de artículos y han sido tenidas en cuenta en el diseño, son por ejemplo evitar un marcado que obligue al autor a utilizar un formato específico para las referencias. En el caso de la sección referencias del artículo, la decisión se basa en que las revistas tienen diferentes políticas publicación que incluyen diversos estilos de citación. La no inclusión de información sobre la revista en la cabecera es otra de las decisiones que pautan el diseño de este *Tag Set*, pues aunque el autor estará creando y remitiendo su artículo a una revista o revistas, no se han incluido opciones para incorporar información específica sobre la revista.

Los módulos que definen este *Tag Set*, sus nombres y funciones son los que aparecen en la Tabla 9.

Tabla 9. Módulos específicos de *Article Authoring DTD*

Fuente: elaboración propia.

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	JATS- articleauthoring1. dtd	<i>Article Authoring DTD module</i>	Módulo principal, o módulo de base del <i>Tag Set</i> : en él se habilitan y deshabilitan los demás módulos específicos, todos subordinados al módulo de base.
<i>Module to name DTD- Specific Modules</i>	JATS- articleauthcusto m-modules1.ent	<i>Article Authoring DTD module of Modules</i>	Este módulo nombra todos los módulos creados expresamente para esta DTD, los cuales no están presentes en la <i>JATS DTD Suite</i> . Declara las entidades de parámetros externas para todos los módulos específicos utilizados por la DTD de autoría. Los módulos solo se declaran en este módulo; se invocan desde el módulo principal del <i>Tag Set</i> .
<i>Customization modules</i>	JATS- articleauthcusto m-classes1.ent	<i>Article Authoring DTD Over-ride Classes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>JATS DTD Suite</i> para las clases de elementos. Como las entidades de parámetro deben ser declaradas antes de ser utilizadas, este módulo es invocado en la DTD de base antes que los módulos de contenido que declaran los elementos, y antes del módulo de clases predeterminadas.
	JATS- articleauthcusto m-mixes1.ent	<i>Article Authoring DTD Over-ride Mixes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>JATS DTD Suite</i> para las mezclas de elementos de propósito general. Este módulo se invoca antes que Suite Default Element Mixes Module en la DTD de base.
	JATS- articleauthcusto m-models1.ent	<i>Article Authoring DTD Over-ride Content and Attribute Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de la <i>JATS DTD Suite</i> . En este módulo, los grupos OR dentro de los modelos utilizan mezclas o clases en lugar de utilizar elementos directamente.

En la versión 1.1 hay dos versiones de la DTD de Autoría: Autoría que utiliza tablas XHTML con MathML 2.0 y Autoría que utiliza tablas XHTML con MathML 3.0. Cada versión de la DTD se corresponde con un fichero ZIP en el FTP, existiendo dos ficheros separados que contienen cada uno una variante de la DTD de Autoría, debiendo el autor elegir una de esas DTD en base a sus requisitos para tablas y MathML.



La última versión oficial de JATS *Article Authoring Tag Set* (versión 1.1, 2015) está disponible en Internet para su descarga en: <ftp://ftp.ncbi.nlm.nih.gov/pub/jats/articleauthoring/1.1/>

#### 4.4. Estructura del marcado en cada *Tag Set*

Cada uno de los tres *Tag Sets* define **contenidos de revista**, entendiéndose por contenidos no solamente el típico artículo de investigación, sino que se incluye en este concepto a todos los componentes de nivel superior en una revista: un artículo de investigación, un artículo de revisión, un comentario sobre un libro o un producto, una carta al editor. El elemento raíz, común a los tres modelos es <article>. Si bien los tres *Tag Sets* se basan en *JATS DTD Suite*, cada uno tiene su nivel de profundidad, diferente flexibilidad respecto a las partes estructurales que utiliza cada modelo, además de diferencias en el número y el uso de los elementos y atributos de la *Suite*, la obligatoriedad o no de los mismos y el orden en el que deben aparecer. Cada *Tag Set* puede incluir o no una parte estructural, y esa parte puede ser requerida (obligatoria) u opcional (Tabla 10).

Tabla 10. Partes estructurales de un artículo con respecto a los tres *Tag Sets* de JATS  
Fuente: elaboración propia.

Partes estructurales del artículo	Conjunto de etiquetas ( <i>Tag Set</i> )		
	Archivo e Intercambio	Publicación	Autoría
1. Parte inicial	Si (requerido)	Si (requerido)	Si (requerido)
2. Cuerpo	Si (opcional)	Si (opcional)	Si (requerido)
3. Parte final	Si (opcional)	Si (opcional)	Si (opcional)
4. Material accesorio	Si (opcional)	Si (opcional)	No requerido
5. Respuestas	Si (opcional)	Si (opcional)	No requerido
6. Artículos subordinados	Si (opcional)	Si (opcional)	No requerido

**Parte inicial / Front matter.** La parte inicial contiene los metadatos del artículo (información de la cabecera del artículo). Estos son el título del artículo, la revista en la que aparece, la fecha y el número de la revista, la declaración de derechos de autor, etc. Tanto los metadatos a nivel de artículo como los metadatos a nivel revista y número están contemplados en los conjuntos de archivo e intercambio y publicación con los elementos <article-meta> y <journal-meta> respectivamente. El conjunto de etiquetas de autoría no incluye el elemento <journal-meta> en la cabecera.

**Cuerpo / Body of the article.** El cuerpo del artículo está compuesto por el texto principal y el contenido gráfico del artículo. Consiste por lo general en párrafos y secciones, que pueden o no contener figuras, tablas, etc. Para los *Tag Sets* de archivo e intercambio y publicación, el cuerpo del artículo es una parte opcional en el marcado de la estructura para permitir a aquellos repositorios que solo mantienen información de la cabecera del artículo la posibilidad de no etiquetar el texto principal.


**Parte final / Back matter.** El artículo puede contener también información que es agregada al texto principal, como un glosario, apéndices o una lista de las referencias citadas en el texto principal.

**Material accesorio / Floating material.** Esta parte está constituida por los objetos que no forman parte del hilo narrativo del texto principal o de la parte final, y que el editor puede optar por

colocar juntos (ejemplo: tablas, figuras, etc.). JATS les asigna el elemento contenedor <floats-group> para facilitar su procesamiento.

A continuación de las cuatro partes mencionadas anteriormente, puede haber una o más **respuestas** al artículo (<response>), o uno o más **artículos subordinados** (<sub-article>).

Las diferencias en las formas de estructurar el marcado por los diferentes *Tag Sets* se pueden apreciar fácilmente comparando los diagramas de jerarquía (*Document hierarchy diagrams*) disponibles en las librerías de etiquetas (*Tag Libraries*) del sitio NLM JATS.

La simbología que utilizan los diagramas que presentamos a continuación y que reflejan gráficamente lo estipulado en la DTD es la siguiente: todos los nombres de los elementos aparecen encerrados en un recuadro. La estructura jerárquica se representa con líneas que unen los diferentes elementos. Cuando la secuencia de un grupo de elementos es requerida, las líneas que los unen forman ángulos rectos, cuando no es requerida los ángulos que forman las líneas no son rectos en el diagrama. Cuando el nombre del elemento aparece solo encerrado en el recuadro, significa que el elemento es requerido. Si el nombre del elemento es precedido por un signo de ?, ese elemento es opcional (ocurrencia 0 o 1). Si el nombre del elemento es precedido por un signo de +, ese elemento es requerido, repetible y su ocurrencia puede ser 1 o más. Si el nombre del elemento es precedido por un asterisco, ese elemento es opcional pero repetible, pudiendo aparecer 0, 1 o más veces. Si luego del nombre del elemento aparece el símbolo ~, significa que el elemento posee atributos. El dibujo  significa que el elemento incluye texto, números y caracteres especiales.

Luego del elemento raíz <article>, común a los tres *Tag Sets*, el siguiente nivel en la estructura es la misma para los *Tag Sets Archiving and Interchange* y *Publishing* pero es diferente para *Authoring*, como se puede apreciar en la Figura 7.

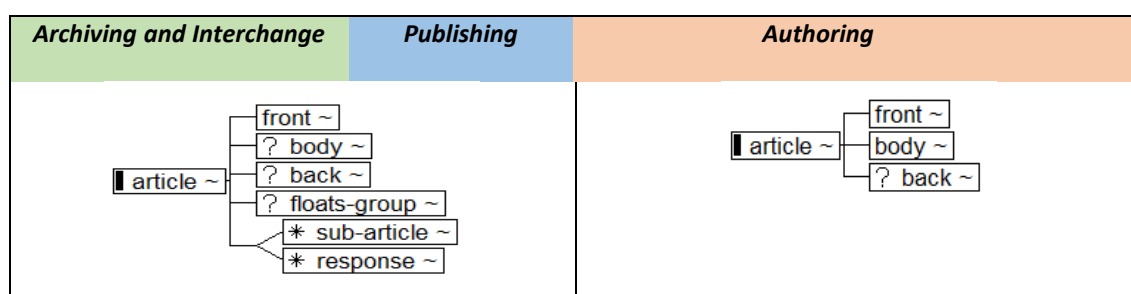


Figura 7. Comparación de las estructuras del elemento raíz en los tres *Tag Sets*.

Fuente: *Document hierarchy diagrams*<sup>4</sup> en las respectivas *Tag libraries*.

En el siguiente nivel de profundidad de la estructura de marcado, y para los elementos que son comunes a los tres *Tag Sets*: <front>, <body> y <back>, se puede apreciar que para *Authoring* solo <back> es opcional, mientras que para *Archiving and Interchange* y *Publishing* solo <front> es obligatorio.

El elemento <front> es un elemento obligatorio para los tres *Tag Sets*, pero tiene diferentes estructuras de marcado en los tres (Figura 8).

<sup>4</sup> *Document Hierarchy diagrams* <http://jats.nlm.nih.gov/publishing/tag-library/1.1d3/chapter/nfd.html>

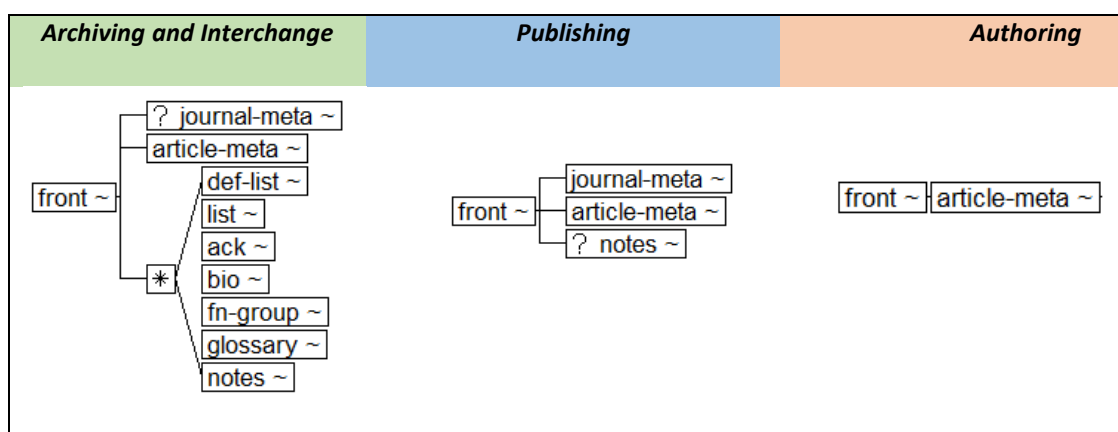


Figura 8. Comparación de las estructuras del elemento <front> en los tres Tag Sets.

Fuente: *Document hierarchy diagrams* en las respectivas *Tag Libraries*.

El elemento <body> también tiene algunas diferencias estructurales según el Tag Set (Figura 9).

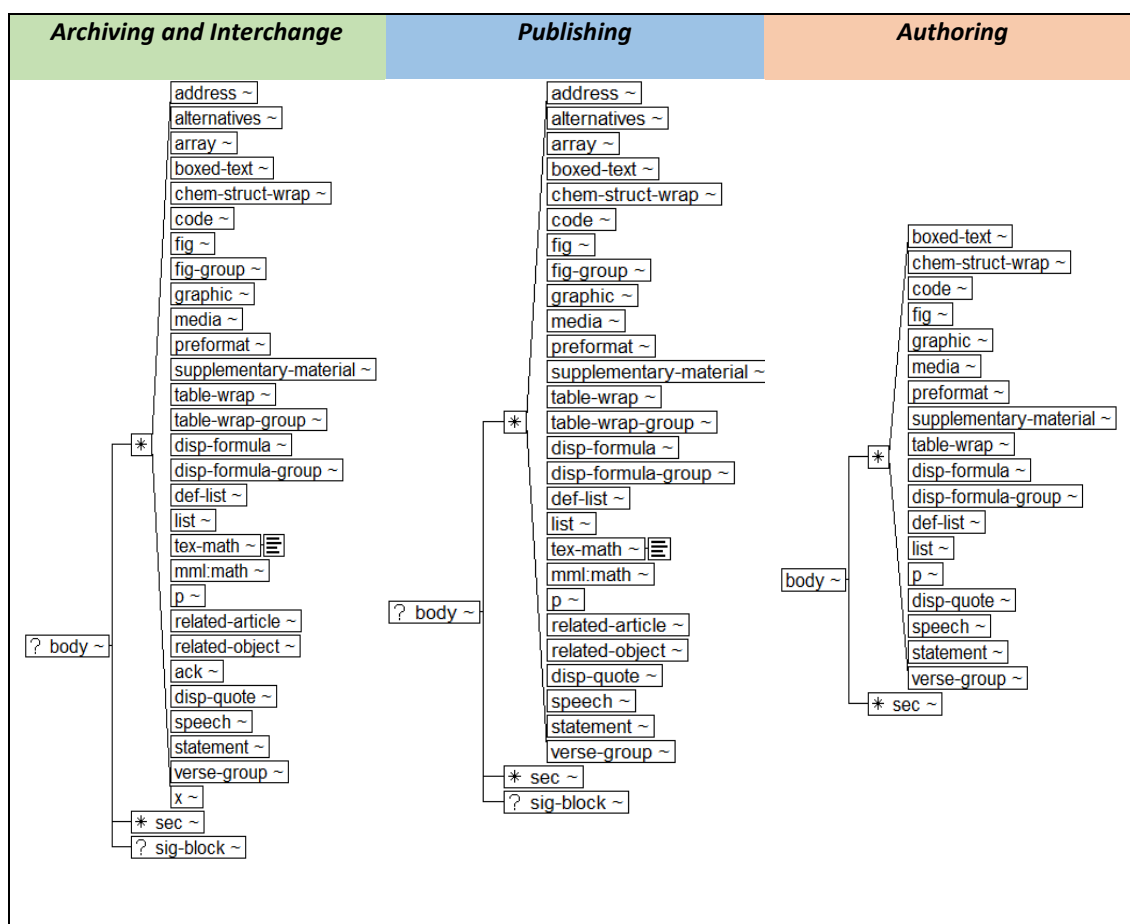


Figura 9. Comparación de las estructuras del elemento <body> en los Tag Sets

Fuente: *Document hierarchy diagrams* en las respectivas *Tag Libraries*.

El elemento <back> tiene la misma estructura en *Archiving and Interchange* y *Publishing*, y es diferente para *Authoring* (Figura 10).

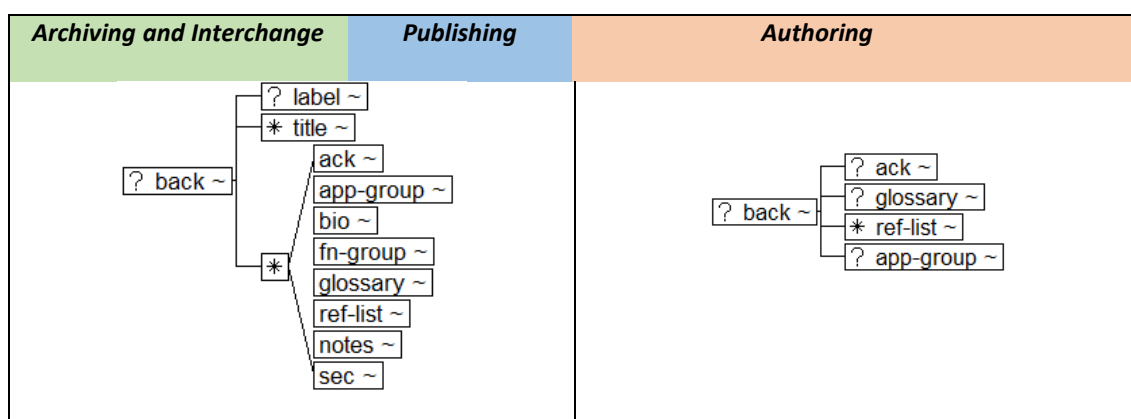


Figura 10. Comparación de las estructuras del elemento `<back>` en los Tag Sets

Fuente: *Document hierarchy diagrams* en las respectivas Tag Libraries.

Al mismo nivel que los elementos `<front>`, `<body>` y `<back>`, pero solo utilizados en los Tag Sets de Archivo e Intercambio y en el de Publicación, están los elementos `<floats-group>`, `<response>` y `<sub-article>`. La estructura de estos elementos es la que se muestra a continuación (Figura 11), siendo la misma para los dos Tag Sets.

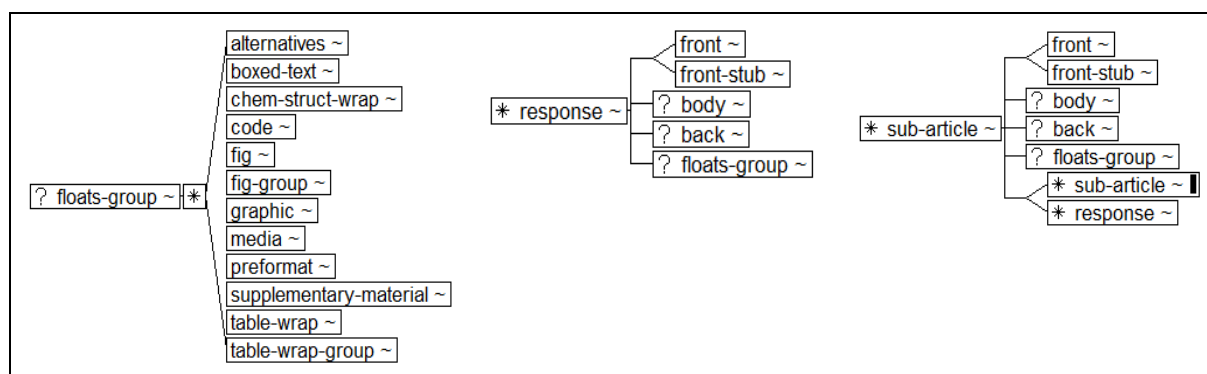


Figura 11. Estructuras de los elementos `<floats-group>`, `<response>` y `<sub-article>` en los Tag Sets Archivo e Intercambio y Publicación

Fuente: *Document hierarchy diagrams* en las respectivas Tag Libraries.

Se pueden dar muchos usos diferentes a los conjuntos de etiquetas, por lo cual además de una estructura estándar, para que el marcado sea coherente se hace necesaria una práctica de marcado en común (*Common Tagging Practice*). En el sitio de NLM JATS existe un espacio para explicar las mejores prácticas de marcado<sup>5</sup>, que ofrece asesoramiento y alternativas sobre cómo proceder con el marcado poniendo algunos ejemplos.

Como ya se ha mencionado anteriormente, JATS fue diseñado deliberadamente para ser un estándar lo suficientemente flexible para que el mayor número posible de editoriales pudieran migrar fácilmente desde sus DTD propias a JATS. En la práctica, esto significa que un mismo documento puede ser marcado de varias formas por distintas editoriales (y a veces con variaciones dentro de los contenidos de una misma editorial) y continuar siendo un archivo JATS XML válido. La preocupación actual de la comunidad de usuarios JATS es que todavía es demasiado permisivo para servir

<sup>5</sup> Mejores prácticas de marcado para *Publishing Tag Set*: <http://jats.nlm.nih.gov/publishing/tag-library/1.1/chapter/tag-practice.html> y Mejores prácticas de marcado para *Archival and Interchange Tag Set*: <http://jats.nlm.nih.gov/archiving/tag-library/1.1/chapter/tag-practice.html>

eficazmente a la reutilización del contenido. Eso es un problema para los sistemas automatizados que reutilizan posteriormente la información, ya que éstos deben identificar con precisión los contenidos de todos los editores y para ello el contenido debe ser marcado de una manera consistente y predecible. Con esa inquietud en mente, en la conferencia anual de JATS de 2014 (JATSSCon 2014) se formó un grupo de trabajo compuesto por representantes de editoriales y proveedores de servicios de procesamiento de contenidos, llamado JATS4R (*JATS For Reuse*), con el objetivo de definir directrices sobre las mejores prácticas de marcado<sup>6</sup> y desarrollar herramientas de validación<sup>7</sup> para ayudar a las editoriales a identificar fácilmente si su contenido es compatible con las prácticas orientadas a la reutilización. El grupo continua trabajando y ha iniciado un proceso para revisar el uso actual de elementos JATS específicos para armonizar su utilización entre las editoriales (Beck y otros, 2015).

---

<sup>6</sup> JATS4R Working Group. General recommendations: <http://jats4r.org/general-recommendations/>

<sup>7</sup> JATS4R Working Group. JATS4R Validator: <http://jats4r.org/validator/>

## CAPÍTULO 5. PERSONALIZACIONES DE JATS

Las personalizaciones de JATS muchas veces son preferidas sobre la versión estándar “tal cual es” (“As-is”) al momento de la implantación. Las editoriales que optan por variaciones del estándar JATS, crean las personalizaciones con el propósito de satisfacer necesidades concretas en sus procesos de publicación o de producción.

La información sobre personalizaciones de JATS que presentamos en este capítulo incluye las extensiones que están disponibles para descarga y uso público en la actualidad. Ha sido recabada de la documentación del estándar en el sitio de NLM, de la Wiki oficial de JATS y de trabajos presentados a las Conferencias anuales *JATS-Con* hasta 2016. Los desarrollos de personalizaciones provienen de NLM, así como también de grupos de trabajo conformados por expertos de diversos proyectos editoriales, que han modificado el estándar en base a requerimientos propios.

Este capítulo se divide en tres secciones. La primera introduce el concepto de personalización, los niveles y las formas implementación de las mismas. La segunda sección se ocupa de las personalizaciones desarrolladas a partir de JATS para contenidos de revista, las cuales incluyen otras funcionalidades que el estándar no contempla en su diseño original. La tercera y última sección presenta personalizaciones de JATS que describen otros tipos de documentos diferentes al contenido de revista.

Algunas personalizaciones son breves y otras más extensas y complejas, pero todas las personalizaciones tratadas en este capítulo se basan en la estructura modular de *JATS* para construir sus esquemas. Las tres personalizaciones de JATS que presentamos en la segunda sección de este capítulo tienen su origen en tres proyectos editoriales diferentes que necesitaron satisfacer requerimientos específicos para sus productos finales y optaron por la creación de un *Tag set* personalizado. Son los casos de la extensión para marcar descripciones taxonómicas en los artículos *TaxPub*, de la extensión *jats-submat* para marcar el material complementario de los artículos y de la reducción *SciELO Publishing Schema* (SPS). SPS utiliza un subconjunto de etiquetas de *JATS Publishing Tag Set* más el estilo propio, para lograr un marcado específico, el cual hace posible marcar información relacionada con indicadores bibliométricos (Indicadores de publicación, Indicadores de colección e Indicadores de citación) de la manera que le resulta más útil al proyecto SciELO para generar sus productos.

Para finalizar, la tercera y última parte del capítulo trata sobre dos extensiones que realizan muchas más modificaciones al modelo JATS: las extensiones utilizadas para describir otros tipos de documentos diferentes a las contribuciones de revistas. Estas son BITS (*Book Interchange Tag Set*) e ISOSTS (*ISO Standards Tag Set*). Tanto BITS como ISOSTS son extensiones de JATS que permiten el marcado de libros y estándares ISO respectivamente. Como menciona Maloney (2011) cuando refiere a los múltiples usos de *JATS Tag Suite* y lo inadecuado que puede resultar su nombre respecto al potencial que tiene para describir otros tipos de documentos: “*One unfortunate aspect of the JATS acronym is that the tag suite is not limited to Journal Articles, but can be used for other types of documents (books, for example).*”

El panorama de las extensiones de JATS continúa diversificándose, tanto así que ciertas extensiones se basan a su vez en otras extensiones, como por ejemplo *Hogrefe Book Tag Set* (HoBoTS), una extensión de BITS codificada en Relax NG que se tratará posteriormente.



## 5.1. Creación de nuevos conjuntos, niveles de personalización y formas de implementación

Las extensiones han dado lugar a nuevos conjuntos de etiquetas o *Tag Sets* personalizados, desarrollados a partir de JATS *Tag Suite*. Penev y otros (2012) opinan en base a su experiencia desarrollando extensiones que JATS DTD ha resultado excepcionalmente bien diseñada para la extensión y que por lo tanto hay relativamente poco que ganar al no tomar ventaja de esto último frente a la opción de documentar y desarrollar herramientas de validación para un perfil no extendido de la DTD, sobre todo si la extensión es fácilmente convertible a JATS DTD.

Los organismos de normalización como NISO, W3C, etc., rara vez desarrollan cambios a los estándares basados en motivos de particulares, o considerando únicamente razones abstractas para hacer un cambio. En la práctica, a menudo sucede que los conjuntos más destacados de personalizaciones y extensiones hechas por terceros a un estándar vigente se convierten en candidatas a ser incorporadas en una versión posterior de ese estándar, y en algunos casos efectivamente pasan a formar parte del estándar (O'Brien y otros, 2011).

O'Brien y otros (2011) realizan una síntesis de los niveles de personalización de un estándar (basado en cuan extensas sean las personalizaciones), y de los métodos de implementación de esas personalizaciones. Definir esos dos aspectos es muy importante, pues un perfil de personalización (*customization profile*) para un conjunto de etiquetas/esquema es una combinación del nivel de personalización y el método de implementación del mismo.

Partiendo del estándar “tal cual es”, los niveles de personalización son cinco:

**Extendido (*Extended*):** es cuando se utiliza un conjunto extendido (*superset*) del estándar, el cual define características adicionales pero opcionales. Cualquier XML que es válido de acuerdo al estándar también será válido para la versión extendida del esquema, aunque lo contrario pueda no ser así.

**Reducido (*Reduced*):** este nivel utiliza un subconjunto (*subset*) del estándar, quitando algunas características de éste o haciéndolas inaccesibles. Un documento XML válido para el estándar puede no ser válido para la versión reducida, pero al contrario sí valida correctamente.

**A medida (*Customized*):** en este nivel, si bien se siguen utilizando los mismos nombres para las etiquetas y sus jerarquías son similares a las del estándar, se presentan cambios que son el resultado de una combinación de extensiones y reducciones. El XML que es válido con el estándar es muy posible que no lo sea con el esquema personalizado, y del mismo modo si intentáramos validar el XML marcado con el esquema personalizado, podría no resultar válido para el estándar.

**Construido a partir de (*Built from*):** en este nivel, se utilizan como punto de partida muchos de los módulos del estándar, pero las modificaciones realizadas al estándar incluyen cambios importantes. Un esquema con este nivel de personalización puede incluir por ejemplo cambios de nombre en las etiquetas, cambios no triviales a la jerarquía de las etiquetas, etc.

**Informado por (*Informed by*):** el estándar puede ser referenciado en el diseño de la aplicación, pero no se utiliza directamente en su implementación. Un esquema a este nivel personalización puede en ocasiones compartir nombres o modelos similares a los del estándar, pero éstos no se definen utilizando los mismos módulos.

Existen además dos enfoques básicos para implementar esos niveles de personalización anteriormente mencionados:

**Sobrescribir (*Overrides*):** con este enfoque, los archivos del conjunto de etiquetas estándar no son modificados directamente, sino que los ficheros originales se dejan intactos. Con este método, las definiciones de las etiquetas personalizadas se encuentran dentro de módulos específicos independientes, que sirven para sobrescribir las versiones estándar de esa misma definición de la etiqueta. Una ventaja de este enfoque es que proporciona una división clara entre la definición personalizada del *Tag Set* y la definición original. Sin embargo, este enfoque requiere que el conjunto de etiquetas original haya sido desarrollado específicamente para permitir esas capacidades de sobrescritura.

**Modificaciones (*Modifications*):** este es un enfoque bastante simple y directo comparado con el anterior. Consiste en editar las definiciones de las etiquetas dentro de los módulos proporcionados en el *Tag Set* estándar.

Con el objetivo de analizar mejor la estructura modular de cada una de las extensiones codificadas en DTD estudiadas en este capítulo, se ha recurrido nuevamente a la modalidad que se utilizó en el capítulo anterior, presentando las características de los módulos de cada extensión en una tabla. Solo los módulos específicos de la respectiva extensión son incluidos, pues los módulos comunes a la *Suite* estarán presentes en todas las extensiones. La Tabla 11 muestra la estructura básica de los módulos para crear nuevos *Tag Sets* e incluye información genérica sobre los tipos de módulo, cómo se nombran los ficheros de módulo, los módulos y cuáles son las funciones de los mismos.

En el caso de partir de JATS DTD, al ser una estructura modular, los *Tag Sets* personalizados que se desarrollan a partir de extender los *Tag Sets* del estándar, se crean a partir de los siguientes módulos: una DTD propia para la extensión (*Base module*), un módulo para nombrar sus componentes llamado módulo de módulos (*Tag-Set-specific Module of Modules*), además de tantos módulos de sobrescritura (*Class overrides*, *Mix overrides*, *Models overrides*) y nuevos módulos como sean necesarios (*New models*). Estos últimos son los que introducen los elementos específicos del nuevo *Tag Set*.

Tabla 11. Estructura de los módulos para creación de nuevos *Tag Sets*

Fuente: Elaboración propia en base a *How to make new tag sets* <<http://jats.nlm.nih.gov/publishing/tag-library/1.0/?chap=pub-implementor>>

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	(nombre de la extensión abreviado).dtd	(Nombre completo de la extensión) <i>DTD Module</i>	Es el módulo DTD de base del nuevo <i>Tag Set</i> . Como mínimo, este módulo declara el elemento de nivel superior (ej.: artículo, libro o informe) y otros elementos estructurales que sean únicos para el nuevo tipo de documento.
<i>Tag-Set-specific Module of Modules</i>	(nombre de la extensión abreviado)-modules.ent	(Nombre completo de la extensión) <i>DTD-Specific Modules</i>	Módulo para nombrar todos los nuevos módulos creados expresamente para el nuevo <i>Tag Set</i> .
<i>Class overrides</i>	(nombre de la extensión abreviado)-classes.ent	(Nombre completo de la extensión) <i>Customize Classes Module</i>	Aporta sobrescritura específica para las clases de elementos por defecto de la <i>Suite</i> .

<i>Mix overrides</i>	(nombre de la extensión abreviado)-mixes.ent	(Nombre completo de la extensión) <i>Customize Mixes Module</i>	Aporta sobrescritura específica para las mezclas de clases por defecto de la <i>Suite</i> .
<i>Models overrides</i>	(nombre de la extensión abreviado)-models.ent	(Nombre completo de la extensión) <i>Customize Content and Attributes Module</i>	Aporta sobrescritura específica para los modelos de contenido en los módulos de la <i>Suite</i> (utilizando las entidades de parámetro "-elements" y "-models").
<i>New Models</i>	(nombre de la extensión abreviado)-(nombre del nuevo modelo).ent	(Nombre completo de la extensión) Nombre del Nuevo Módulo	Contiene los nuevos elementos específicos del nuevo <i>Tag Set</i> . Por ejemplo, un nuevo <i>Tag Set</i> para marcar libros añadirá elementos que corresponden a metadatos específicos para libros.

## 5.2. Personalizaciones para artículos

### 5.2.1. TaxPub XML schema

Este esquema es una extensión XML de NLM/NCBI *Journal Publishing DTD*, v 3.0. La versión actual es la 0.5 de junio de 2011. TaxPub fue creado con el objeto de proporcionar elementos de marcado específico para la codificación de artículos que contengan tratamientos taxonómicos, en el ámbito de la sistemática biológica. Según Penev y otros (2012), la idea central del esquema es delimitar descripciones de taxón<sup>8</sup> dentro de un artículo, además de describir varios sub-elementos dentro de un tratamiento taxonómico con la intención de utilizar esas porciones individuales de información para diversos fines, el más importante de los cuales es la recuperación de la descripción del taxón y sus secciones por separado del texto del artículo. De esta manera, los subelementos de la descripción pueden ser cosechados por (o exportados a) diversos agregadores de información sobre biodiversidad.

El "tratamiento taxonómico" es el foco de TaxPub. Éste término refiere a la descripción formal de un taxón, incluyendo secciones sobre nomenclatura, características morfológicas, comportamiento, ecología, distribución y especímenes examinados. La extensión TaxPub provee un modelo para esas características de un tratamiento taxonómico, proporcionando dentro de un *namespace* con el prefijo "tp" un elemento <tp:taxon-treatment> con un elemento requerido llamado <tp:nomenclature> altamente estructurado que contiene los datos esenciales sobre la especie, y un elemento <tp:treatment-sec> para otras sub-secciones para las que un atributo de tipo treatment-sec-type proporciona la semántica específica. La intención de los desarrolladores de esta extensión es que los usuarios empleen términos de vocabularios controlados externos, como por ejemplo DarwinCore<sup>9</sup> para suministrar esa semántica específica.

TaxPub introduce además en la mayoría de los elementos de bloque de *Journal Publishing DTD* las siguientes etiquetas: <tp:taxon-name> para marcar los nombres científicos de los organismos, <tp:material-citation> para marcar las referencias a materiales relacionados a los especímenes, y <tp:descriptive-statement> para marcar la descripción de las características físicas de los organismos.

<sup>8</sup> Taxón: 1. m. Biol. Cada una de las subdivisiones de la clasificación biológica, desde la especie, que se toma como unidad, hasta el filo o tipo de organización. Fuente: Diccionario RAE 23ª. Ed.

<sup>9</sup> Biodiversity Information Standards - Darwin Core <http://rs.tdwg.org/dwc/>

Más allá de estas incorporaciones, TaxPub se basa en los elementos de *Journal Publishing DTD* para todas sus demás características.

En la Tabla 12 se puede apreciar claramente la estructura modular de base JATS, así como también el nuevo módulo con los elementos específicos incorporados por TaxPub: *taxpubcustom-elements.ent*.

Tabla 12. Módulos específicos de TaxPub DTD (v0.5 Jun. 2011)

Fuente: elaboración propia

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	tax-treatment-NS0.dtd	<i>Taxonomic Treatment Publishing DTD module</i>	Módulo principal, o módulo de base. Todos los demás módulos específicos están subordinados al módulo de base. Su propósito es declarar las entidades de parámetro que se usan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>Journal Publishing DTD</i> .
<i>Module to name DTD-specific modules</i>	taxpubcustom-modules-NS0.ent	<i>Taxonomic Treatment Module of Modules</i>	Este módulo nombra los módulos creados expresamente para esta DTD, y declara las entidades de parámetro externas para todos los módulos específicos utilizados por esta DTD. Los módulos solo se declaran en este módulo, y se invocan desde el módulo principal de la DTD.
<i>Customization modules</i>	taxpubcustom-classes-NS0.ent	<i>Taxonomic Treatment Publishing Customize Classes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>Publishing DTD</i> para las clases de elementos.
	taxpubcustom-classes3-NS0.ent	<i>Taxonomic Treatment Publishing Mixes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>Publishing DTD</i> para las mezclas de elementos de propósito general.
	taxpubcustom-models-NS0.ent	<i>Taxonomic Treatment Publishing Content and Attributes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>Publishing DTD</i> .
<i>New models</i>	taxpubcustom-elements.ent	<i>Taxonomic Treatment Publishing Elements Module</i>	Declara los elementos que se utilizan para describir los tratamientos taxonómicos, siguiendo las convenciones existentes, altamente formalizadas.

La extensión está disponible en Internet para su descarga en: <https://github.com/plazi/TaxPub> . No se provee documentación adicional.

### 5.2.2. Jats-supmat

El objeto de esta extensión es establecer un conjunto básico de información requerida, con el objeto de asegurar que el material complementario del artículo posea una identificación, un enlace al artículo, informe sobre el nivel de importancia, e incluya información sobre el formato (Gutzman y Tryka, 2013). La versión actual es 0.1 y no es una versión final, sino que es considerada un borrador de trabajo.

A raíz de una encuesta sobre cómo los editores trabajan con materiales complementarios, NISO y *National Federation of Advanced Information Systems* (NFAIS), formaron un grupo de trabajo para identificar y abordar cuestiones relacionadas con el material complementario de los artículos. Los resultados del trabajo de dicho grupo fueron publicados como un conjunto de mejores prácticas recomendadas en enero de 2013: "*Recommended Practices for Online Supplemental Journal Article Materials*" (NISO RP-2013). *JATS-supmats* es una extensión de JATS que incorpora la recomendación NISO RP-2013 en su diseño.

En la descripción de material complementario, la NISO RP define una serie de valores de metadatos recomendados y opcionales en los cuales se basan los autores de esta extensión para su diseño. Guzman y Tryka (2013) han creado la extensión a partir de una nueva DTD de base que llamaron NISO-supmat-extension-rec0.dtd , la cual invoca a los ficheros de la *Journal Publishing DTD* y además los siguientes ficheros personalizados con los cambios y agregados de la extensión: NISO-supmatcustom-classes.ent, NISO-supmatcustom-elements.ent, NISO-supmatcustom-mixes.ent, NISO-supmatcustom-models.ent, NISO-supmatcustom-modules.ent.

En la Tabla 13 se puede apreciar la estructura modular de base JATS utilizada para la creación del nuevo *Tag Set*, así como también el nuevo módulo incorporado por JATS-supmat: NISO-supmatcustom-elements.ent

Tabla 13. Módulos específicos de JATS-supmat DTD (v0.1 20130503)

Fuente: elaboración propia.

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	NISO-supmat-extension-rec0.dtd	<i>Journal Publishing SupMat extension DTD module</i>	Es el nuevo módulo principal, o módulo de base. Todos los demás módulos específicos están subordinados al módulo de base.
<i>Module to name DTD-specific modules</i>	NISO-supmatcustom-elements.ent	<i>JATS Journal Publishing SupMat Elements Module</i>	Este módulo nombra los módulos creados expresamente para esta DTD, y declara las entidades de parámetros externas para todos los módulos específicos utilizados por esta DTD.
	NISO-supmatcustom-classes.ent	<i>JATS Journal Publishing SupMat Customize Class Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>Publishing DTD</i> para las clases de elementos.

Customization modules	NISO-supmatcustom-mixes.ent	JATS Journal Publishing SupMat Customize Mixes Module	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>Publishing</i> DTD para las mezclas de elementos de propósito general.
	NISO-supmatcustom-models.ent	JATS Journal Publishing SupMat Customize Content and Attributes Module	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de <i>Publishing</i> DTD.
New Models	NISO-supmatcustom-elements.ent	JATS Journal Publishing SupMat Elements Module	Declara los elementos que se utilizan para describir el material suplementario.

La extensión está disponible para su descarga en: <https://github.com/Klortho/jats-supmat>

### 5.2.3. SciELO Publishing Schema (SPS)

SciELO PS se basa en un subconjunto de *JATS Publishing 1.0* (Fonseca, 2016). Los elementos y atributos XML utilizados en SPS son los mismos que define JATS, pero SPS usa algunos elementos de *JATS Publishing Tag Set* con estilo propio (el estilo SciELO), sin que esto implique modificar dichos elementos. Esto significa que SciELO PS no modifica los elementos que toma de *JATS Publishing Tag Set*, sino que define la forma en que éstos deben ser utilizados en el esquema. Según Fonseca (2016), esa forma de utilización incluye vocabularios controlados para los valores de los atributos de los elementos, restricciones de uso de elementos o atributos, etc., aclarando además que no todos los elementos listados en la documentación de SPS representan necesariamente una regla en el estilo SciELO. Las reglas del estilo SciELO son validadas con un esquema externo Schematron, y esa validación se realiza después de la validación del XML de los artículos con *JATS Publishing DTD* versión 1.0.

El estilo SciELO es implementado en Schematron, un lenguaje de marcado XML (más específicamente un lenguaje de validación), que en el caso del estilo SciELO es utilizado conjuntamente con otros lenguajes de esquema (DTD). Resulta en general muy útil como método de comprobación para extender validaciones basadas en DTD, RELAX NG o XML Schema, pues puede imponer restricciones que no pueden expresarse en los otros lenguajes de esquemas.

Entre los servicios de la plataforma SciELO se encuentran la estructuración de texto en lenguaje XML para el almacenamiento en bases de datos, la extracción de metadatos, los cálculos bibliométricos y la interoperabilidad con otros recursos de Internet. Es para cumplir con requerimientos específicos que SciELO necesita marcar datos esenciales para las funciones de indexación: el marcado de la afiliación institucional de todos los autores (necesaria para generar los indicadores de producción científica por institución y sus principales unidades: Universidad, Facultad, etc.), el marcado de los elementos que componen las referencias de los documentos citados en los artículos (indispensables para generar los índices bibliométricos de SciELO), y el marcado que permita la identificación de las fuentes de financiamiento de la investigación.

La versión actual es la 1.4 de marzo de 2016. La especificación SPS, una reducción basada en *JATS Journal Publishing DTD* (versión 1.0) es totalmente compatible con el estándar y con PMC DTD.



La personalización SPS está disponible para su descarga en [https://github.com/scieloorg/scielo\\_publishing\\_schema](https://github.com/scieloorg/scielo_publishing_schema) y el fichero XSD en <http://static.scielo.org/sps/schema/SciELO-journalpublishing1.xsd> ; la documentación sobre su uso puede encontrarse en *SciELO Publishing Schema - Guía de uso dos elementos e atributos para documentos em XML* [http://docs.scielo.org/projects/scielo-publishing-schema/pt\\_BR/latest/](http://docs.scielo.org/projects/scielo-publishing-schema/pt_BR/latest/)

### 5.3. Personalizaciones para describir otros tipos de documentos diferentes al artículo de revista

#### 5.3.1. Book Interchange Tag Suite (BITS)

El agregado de módulos adicionales a *NISO JATS Suite* dio lugar a *Book Interchange Tag Suite* (BITS). Si bien BITS es una extensión realizada en base a los módulos JATS, no fue parte del diseño original de JATS y no es un estándar NISO. Por lo tanto, BITS no es administrado ni actualizado por NISO, sino por el Centro Nacional de Información Biotecnológica (NCBI) de la Biblioteca Nacional de Medicina de Estados Unidos (NLM).

La *Suite* de módulos de BITS DTD fue escrita como base para la publicación, intercambio y archivo, con la intención de que DTD para fines específicos se puedan desarrollar en base a esos módulos. Su cometido es proporcionar un conjunto de módulos de esquema XML, los cuales definen los elementos y atributos para describir el contenido textual y gráfico de libros, y también de sus partes componentes. Además, provee los metadatos a nivel de colección para series de libros y colecciones de libros, cuando una parte del libro está asociada con una o más de dichas colecciones.

La *Suite* tiene como antecedente a NCBI Book DTD, y Beck (2013) opina que realmente es una mejora en relación a su antecesor, pues gestiona mejor la estructura de los libros y contempla objetos específicos de los libros que las editoriales habían estado solicitando.

*BITS is a real improvement over the old NCBI Book DTD. The structure of the book is handled better and there are more book-specific objects that book publishers have been requesting [...] BITS was intended to be a fresh start at a general book model that is familiar to JATS users and not simply an update to the NCBI Book DTD* (Beck, 2013).

Lo que impulsó el desarrollo de BITS fue el convencimiento de que un modelo para libros basado en XML sería útil a una amplia variedad de editores de libros académicos y sobre todo, pero no exclusivamente, a las editoriales que ya utilizan uno de los modelos JATS para artículos de revista y buscan un modelo compatible para sus libros. Por esa razón es que el diseño de BITS DTD fue desarrollado como una extensión de ANSI/NISO JATS Z39.96-2012 *Journal Archiving and Interchange Tag Set*, siendo la extensión la que provee los agregados para describir libros, partes de libros (capítulos), e información sobre la inclusión de libros o las partes de los libros en series monográficas. BITS cumple además con el propósito de ser útil para el intercambio de libros y partes de libros, al proporcionar un formato común con el que las editoriales y los archivos pueden intercambiar esos contenidos.

Cuando el grupo de trabajo planteó el modelo para BITS, acordaron que debía contemplar libros completos y también componentes de libros (un capítulo, parte o módulo), que debía definir tanto el material nuevo como el de los libros ya existentes, y ser capaz de describir volúmenes y series de libros. El alcance de BITS incluye monografías técnicas, informes gubernamentales, series de monografías de varios volúmenes, libros de referencia, actas de congresos, enciclopedias, así como también contenido tradicionalmente alojado en la web, tales como *Gene Reviews*, donde cada registro

puede considerarse equivalente a un capítulo de un libro. Los atlas y las guías de campo no se consideran fuera de su alcance, pero pueden necesitar metadatos semánticos adicionales para poder ser representados completamente (Beck, 2013).

El diseño de BITS se basa al igual que JATS, en una DTD modular cuyos módulos se pueden combinar para obtener diferentes conjuntos de etiquetas. BITS DTD se construyó utilizando los módulos de JATS y añadiendo módulos adicionales que son propios de BITS. Si JATS ya nombra una estructura y esa estructura aparece en el contenido de un libro, entonces el nombre JATS, y en lo posible el modelo JATS ha sido utilizado “tal cual es”. Al igual que en JATS, los módulos son archivos separados que en conjunto definen todas las estructuras de elementos como tablas, fórmulas matemáticas y químicas, párrafos, secciones, figuras, notas al pie, y los elementos de las referencias, así como también los atributos y entidades de la *Suite*. El *Tag Set* se basa en la versión 1.0 de *Journal Archiving and Interchange Tag Set* e incluye sus capacidades multi-idioma. Al igual que JATS, está disponible en DTD, Relax NG y W3C XML Schema.

Una de las decisiones básicas que adoptó el grupo de trabajo cuando diseñó BITS fue que esta extensión debería ser capaz de representar el libro como un único documento XML, o sea con un elemento raíz (<book>) o bien compuesto de un conjunto de capítulos (<book-part>). Esto está en línea con el modelo de artículo de JATS, el cual no describe números ni revistas, pero incluye suficiente información en cada artículo para permitir que un sistema de publicación pueda crearlos a partir de una colección de artículos. Es por ello que en BITS hay dos elementos de nivel superior: libro: <book> para contener todo un libro y <book-part-wrap> para contener una parte de libro, por ejemplo un capítulo o un módulo que se considere como unidad.

El elemento <book-part-wrap> es para describir una o más partes de libros <book-part>. Para construir un libro a partir de una colección de partes, cada <book-part> debe incluir o tener asociada información sobre los metadatos a nivel de libro, como cada artículo JATS incluye información sobre los metadatos a nivel revista (<journal-meta>) y número de revista (<article-meta>). El elemento contenedor <book-part-wrap> incluye los elementos <book-meta> y <book-part> (repetible). Eso está estructurado así para mantener los metadatos a nivel de libro fuera del contenido de <book-part>.

Se hacen necesarias en esta extensión nuevas estructuras específicas para libros (*Book-specific objects*), por lo cual BITS incorpora dos nuevos elementos a nivel de objeto, tanto para libro como para partes de libro, las cuales permiten marcado de específico de los objetos tabla de contenido e índice: <toc> e <index> respectivamente.

BITS DTD está constituida de diferentes tipos de módulos específicos (Tabla 14), que se han desarrollado a partir de los módulos de *JATS DTD*, realizando cambios en las declaraciones de módulos de la *Suite* mediante la sobrescritura de los contenidos de sus entidades de parámetro. Esa sobrescritura está definida en los tres módulos personalizados de BITS DTD, BITS-bookcustom-classes.ent, BITS-bookcustom-mixes.ent y BITS-bookcustom-models.ent, los cuales son invocados desde el módulo principal (*Base module*): BITS-book1.dtd.

Tabla 14. Módulos específicos de Book Interchange DTD (v1.0 Dec. 2013)

Fuente: elaboración propia en base a la documentación de BITS.

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	BITS-book1.dtd	<i>Book Interchange DTD module</i>	Módulo principal, o módulo de base del <i>Tag Set</i> . Todos los demás módulos específicos están subordinados al módulo de base. Su propósito es declarar las entidades de parámetro que se usan para sobrescribir los modelos de

			contenido, partes de modelos de contenido, o listas de atributos de <i>BITS DTD Suite</i> .
<i>Module to name DTD-specific modules</i>	BITS-bookcustom-modules1.ent	<i>Book Interchange DTD-Specific Modules</i>	Este módulo nombra los módulos creados expresamente para esta DTD, es decir, los que no están presentes como parte de los fragmentos de BITS DTD o los fragmentos de NISO JATS DTD. Declara las entidades de parámetro externas para todos los módulos específicos utilizados por BITS DTD. Los módulos solo se declaran en este módulo, y se invocan desde el módulo principal del <i>Tag Set</i> .
<i>Customization modules</i>	BITS-bookcustom-classes1.ent	<i>Book Interchange DTD Customize Classes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de JATS DTD Suite para las clases de elementos. Como las entidades de parámetro deben ser declaradas antes de ser utilizadas, este módulo es invocado en la DTD de base antes que los módulos de contenido que declaran los elementos, y antes del módulo de clases predeterminadas.
	BITS-bookcustom-mixes1.ent	<i>Book Interchange DTD Customize Mixes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de <i>JATS DTD Suite</i> para las mezclas de elementos de propósito general.
	BITS-bookcustom-models1.ent	<i>Book Interchange DTD Customize Content and Attributes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de la <i>JATS DTD Suite</i> . En este módulo, los grupos OR dentro de los modelos utilizan mezclas o clases en lugar de utilizar elementos directamente.
<i>New models</i>	BITS-bookmeta1.ent	<i>BITS Book Metadata Elements Module</i>	Define los elementos, atributos y entidades de parámetro que son compartidos entre todas las DTD BITS pero no entre las DTD JATS.
	BITS-book-part1.ent	<i>BITS Book Component (Book-Part) Elements Module</i>	Declara los elementos del libro a nivel de componentes, como los elementos específicos de capítulos o específicos de otras partes componentes de un libro.
	BITS-book-part-wrap1.ent	<i>BITS Top-Level Book-Part-Wrapper Module</i>	Declara los elementos del nivel superior para partes de libro <book-part-wrap>
	BITS-index1.ent	<i>BITS Structural Index Elements Module</i>	Define los elementos, atributos y entidades de parámetro específicos para un índice estructural.

	BITS-embedded-index1.ent	<i>BITS Embedded Index Elements Module</i>	Define los términos del índice a ser incorporados en el texto de un documento, con el propósito de poder generar un índice automático mediante programación.
	BITS-toc1.ent	<i>BITS Structural Table Of Contents Module</i>	Define los elementos, atributos y entidades de parámetro específicos para una tabla de contenidos estructural.
	BITS-toc-index-nav1.ent	<i>BITS Structural Navigation Module</i>	Define los elementos utilizados para la navegación en los índices estructurales y tablas de contenido.
	BITS-question-answer1.ent	<i>BITS Question And Answer Module</i>	Define un modelo básico para preguntas y para conjuntos de preguntas y respuestas.

La versión actual de BITS es la 2.0 de febrero de 2016. Está basada en JATS v.1.1. y por ello cuenta con dos variantes de BITS DTD basados en la *Suite*, las cuales son idénticas excepto en lo que respecta modelado de las tablas: BITS DTD con tablas XHTML (BITS-book2.dtd) que incluye el modelo JATS XHTML para las tablas y BITS OASIS (BITS-book-oasis2.dtd) que incluye dos modelos de tablas: JATS XHTML y tablas OASIS CALS Exchange.

La extensión y sus correspondientes versiones (incluidas las librerías de etiquetas) están disponibles para descarga vía FTP: <ftp://ftp.ncbi.nih.gov/pub/jats/extensions/bits/>.

BITS ha sido pensada desde su creación y diseño para poder generar extensiones a partir del uso de un nuevo archivo DTD y nuevos módulos específicos de personalización para redefinir las entidades de parámetros. Como en todos los casos en los que se crean nuevas extensiones, si se alterara uno o más módulos de la *Suite*, se debe cambiar el nombre de la versión y el de todos sus módulos relacionados, para evitar cualquier confusión con la *Suite* original. Este es el caso de *Hogrefe Book Tag Set* (HoBoTS), mencionado en la introducción de este capítulo. HoBoTS es una extensión de BITS creada por el grupo editorial Hogrefe, una editorial europea de libros, revistas, y pruebas en el campo de la psicología y disciplinas afines, para ser utilizada en su flujo de trabajo XML.

El grupo editorial Hogrefe estaba buscando una solución para la creación de productos electrónicos en XML para sus libros. Cuando comenzaron el proyecto estaba disponible BITS versión beta 0.2 y partieron de esa DTD para crear HoBoTS, de la cual además extendieron el modelo de atributos comunes de JATS (*Common Attributes content model*) para contener atributos CSS y RDFa.

Para crear el nuevo esquema convirtieron BITS DTD a Relax NG con una herramienta de conversión denominada Trang<sup>10</sup>, un procedimiento que les resulta fácilmente reproducible cuando aparece una nueva versión de BITS. La razón por la cual utilizaron Relax NG es porque con esta codificación resulta más fácil incluir, ampliar y restringir el esquema, sin la necesidad de modificar el código fuente del que ha sido convertido automáticamente (Imsieke, 2013).

<sup>10</sup> Trang. Multi-format schema converter based on RELAX NG  
<http://www.thaiopensource.com/relaxng/trang.html>

En base a la documentación de la extensión, los cambios que introduce HoBoTS son los siguientes: el valor del atributo `dtd-version` es *1.0-variant Hogrefe Book Tag Set (hobots) 1.0*. Este atributo puede ir acompañado de otro atributo `css:version`. En cuanto a los elementos estructurales las modificaciones están en el elemento `<toc>` de BITS, que puede estar presente en este nuevo esquema aun cuando el documento no contenga una tabla de contenido completa, y en el elemento `<index>`, que puede estar presente aun cuando no tenga ningún contenido y puede llevar el atributo `sort-key`.

Hogrefe ofrece esta personalización gratuitamente a disposición del público. La extensión HoBoTS está disponible para descarga en: <http://hobots.hogrefe.com/schema/hobots.zip>

### 5.3.2. ISO Standards Tag Set (ISOSTS)

ISOSTS es un *Tag Set* desarrollado por la Secretaría Central de *International Standards Organization* (ISO) en 2011. Su versión actual es la 1.1 de julio de 2013. Está basado parcialmente en el conjunto de etiquetas NISO JATS versión 0.4 y ha sido creado como una extensión de JATS *Journal Publishing* DTD. Una característica a destacar de esta extensión es que el nivel superior y ciertos metadatos han sido personalizados, por lo cual no es totalmente compatible con NISO JATS.

JATS es una buena opción para los estándares ISO debido a que cuenta con capacidades para tratar documentos multilingües y eso es de suma importancia para un organismo internacional. Los conjuntos actuales de etiquetas del estándar JATS proporcionan varias estructuras para la codificación de documentos en los que algunos de los metadatos o el texto estén en varios idiomas. El contenido en varios idiomas puede ser manejado usando tres técnicas sencillas: 1) la identificación del idioma mediante el atributo genérico de XML `@xml:lang`, 2) la repetición de algunas estructuras de modo de estar presentes en más de un idioma, y 3) adjuntando algunas estructuras repetibles en un solo elemento contenedor, para indicar que contienen representaciones de un objeto lógico individual en diferentes idiomas (Lapeyre y Usdin, 2011). En ISOSTS, los documentos multilingües son manejados con estructuras en el XML que se pueden repetir y que llevan información sobre el idioma en forma de atributos.

ISOSTS está diseñado para marcar documentos de estándares generados por organizaciones como ISO, *European Committee for Standardization* (CEN), y otros organismos a nivel internacional, regional y nacional que se dedican al desarrollo de estándares. Las estructuras de metadatos necesarios para describir estándares son significativamente diferentes de las necesarias para artículos de revistas, por lo cual ciertos metadatos en este conjunto de etiquetas han sido creados expresamente. Si bien se ha reutilizando parte del conjunto de etiquetas JATS para su aplicación al marcado de estándares, existen cambios importantes a nivel estructural en ISOSTS. La sección de metadatos a nivel de artículo de *Journal Publishing Tag Set* ha sido completamente suprimida y reemplazada con tres bloques de metadatos distintos, diseñados para contener los metadatos ISO (`<iso-meta>`), los metadatos CEN (`<cen-meta>`), o bien los metadatos de organismos nacionales de normalización (`<nat-meta>`).

En ISOSTS el elemento raíz ya no es `<article>`, sino `<standard>`. Muchos de los nuevos elementos son específicos para los metadatos ISO y son hijos del elemento contenedor `<iso-meta>`: `<comm-ref>`, `<compl>`, `<content-language>`, `<doc-ident>`, `<doc-number>`, `<doc-ref>`, `<doc-type>`, `<full>`, `<ics>`, `<intro>`, `<is-proof>`, `<language>`, `<main>`, `<originator>`, `<part-number>`, `<proj-id>`, `<release-date>`, `<release-version>`, `<sdo>`, `<secretariat>`, `<std-ident>`, `<std-ref>`, `<suppl-number>`, `<suppl-type>`, `<suppl-version>`, `<title-wrap>`, `<urn>`, y `<version>`. También se han agregado elementos de marcado semántico para información terminológica: `<tbx:termEntry>` y `<term-display>` se utilizan para codificar la terminología definida en los estándares.

Además, algunos elementos JATS se modifican. Los cambios que se han realizado a algunos de los elementos del vocabulario de base que aporta JATS consisten en modificaciones a sus definiciones formales o a su documentación, para ponerlos al servicio de los fines de ISOSTS. Aquellos cuyas definiciones han sido modificadas son: <front> (para incluir metadatos del tipo de documento standards), <sec> (para incluir <term-sec> y omitir metadatos de nivel *section*), <app> (para incluir <annex-type>), <boxed-text> (para omitir metadatos de nivel *section*), <pub-date> (para especificar contenido ISO 8601-style), <ref> (se utiliza para excluir notas en las citas).

Puesto que el cuerpo de los estándares es muy similar al de la mayoría de los demás materiales técnicos, las estructuras dentro de <body> se han tomado de JATS. ISOSTS las ha adoptado con muy pocos cambios. Utiliza además los mismos mecanismos de extensión y las mismas convenciones para las entidades de parámetro que la DTD de JATS.

En la Tabla 15 se puede apreciar la estructura modular de base JATS y los nuevos módulos incorporados por la extensión ISOSTS (*New models*).

Tabla 15. Módulos específicos de ISO Standards Tag Set DTD (v1.1 Jul. 2013)  
Fuente: elaboración propia

Tipo de módulo	Ficheros de módulo	Nombre del módulo	Función
<i>Base module</i>	ISOSTS.dtd	<i>ISOSTS DTD module</i>	Módulo principal del <i>Tag Set</i> . Es el nuevo módulo de base en lugar de JATS-journalpublishing0.dtd. Todos los demás módulos específicos de la extensión están subordinados a este módulo de base. Su propósito es declarar las entidades de parámetro que se usan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de ISOSTS.
<i>Module to name DTD-specific modules</i>	ISOSTS-modules.ent	<i>ISOSTS Module of Modules</i>	Este módulo nombra los módulos creados expresamente para esta DTD. Declara las entidades de parámetro externas para todos los módulos específicos utilizados por ISOSTS DTD. Los módulos solo se declaran en este módulo, y se invocan desde el módulo principal del <i>Tag Set</i> .
<i>Customization modules</i>	ISOSTS-classes.ent	<i>ISOSTS DTD Customize Classes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de JATS DTD Suite para las clases de elementos.
	ISOSTS-mixes.ent	<i>ISOSTS DTD Customize Mixes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir la opción predeterminada de JATS DTD Suite para las mezclas de elementos de propósito general.
	ISOSTS-models.ent	<i>ISOSTS DTD Custom Content and Attributes Module</i>	Declara las entidades de parámetro que se utilizan para sobrescribir los modelos de contenido, partes de modelos de contenido, o listas de atributos de la JATS DTD Suite.



New Models	ISOSTS-iso-metadata.ent	<i>ISOSTS ISO Metadata Module</i>	Define los metadatos específicos de ISO, concretamente el elemento iso-meta, sus contenidos y entidades de parámetro para la personalización de modelos de contenido y listas de atributos.
	ISOSTS-nat-metadata.ent	<i>ISOSTS National Body Metadata Module</i>	Define elementos de metadatos específicos a los organismos nacionales que puedan trabajar con estándares ISO codificadas con esta DTD. En la versión actual de la DTD, el contenido de <nat-meta> sigue el modelo de <iso-meta>.
	ISOSTS-reg-metadata.ent	<i>ISOSTS Regional Body Metadata Module</i>	Define los elementos específicos para las organizaciones que desarrollan estándares y que se clasifican como regionales. En la versión actual de esta DTD, el contenido de <reg-meta> se basa en el modelo <iso-meta>, el cual se extendió para contemplar los requisitos de CEN.
	ISOSTS-tbx-namespace.ent	<i>ISOSTS TBX Terminology Module</i>	Define un vocabulario compatible en base a un espacio de nombres calificado, TBX (TermBase eXchange) derivado de la norma ISO 30042, utilizado para el marcado de texto y para los términos.
	ISOSTS-tbx-setup.ent	<i>TBX DTD Setup Module</i>	Su función principal es invocar a <i>TBX Terminology Module</i> , además sobrescribe algunas entidades de parámetro estándar.

La extensión está disponible en Internet para su descarga en <http://www.iso.org/schema/isosts/> y su documentación en <http://www.iso.org/schema/isosts/v1.0/doc/>

Cabe destacar que ISOSTS no es un estándar, y existen además varias DTD utilizadas para etiquetar la información de tipo estándar basadas en JATS y otras desarrolladas de forma independiente. En la sección noticias del sitio web de NISO, fechado el 5 de agosto de 2015, se menciona expresamente que la existencia de varias formas de marcar documentos de estándares impide la interoperabilidad entre documentos producidos por diferentes organizaciones. Al ser ISOSTS una de las extensiones de JATS cuya DTD y la documentación correspondiente han sido puestas a disposición en forma abierta, está siendo utilizada actualmente por varios de los miembros de ISO. (NISO, 2015).

Los comités técnicos creados a los efectos de trabajar en la elaboración del futuro estándar, denominado NISO STS, se están reuniendo desde octubre de 2015. Dentro de las decisiones más importantes adoptadas a la fecha por los mencionados comités, se incluyen aspectos de diseño como que NISO STS se basará en el modelo JATS *Publishing* (azul) v. 1.1 y que se adoptarán los modelos de la extensión BITS para índices y tablas de contenido. También tienen planeado que el modelo XML para NISO STS esté disponible en DTD, XSD, y RelaxNG (Wheeler y otros, 2016).

## CAPÍTULO 6. SITUACIÓN ACTUAL DE JATS

En este capítulo se aborda la implantación de JATS como estándar de marcado por parte de diversas editoriales, bibliotecas y archivos digitales de revistas en la actualidad. Muchos de ellos no se limitan a implantar JATS “tal cual es”, sino que han desarrollado sus propias personalizaciones del estándar, o utilizan las personalizaciones que están disponibles en acceso libre. Algunos implementan JATS sin modificaciones, pero no marcan el texto completo sino solo algunos metadatos, o tienen una metodología o estilos de marcado propio, etc. Partiendo de casos concretos de implementación de JATS, pretendemos mostrar algunos ejemplos de la diversidad de emprendimientos que utilizan el estándar o alguna de sus personalizaciones y cómo la versatilidad de su diseño se presta para la consecución de diversos objetivos.

Como menciona Rosenblum (2015), hace 15 años, la opción XML para los editores de revistas requería o bien un proyecto de I+D para crear una DTD, o adoptar la DTD de un determinado proveedor (la editorial), con la posibilidad de que posteriormente pudiera resultar problemática esta decisión si luego se cambiaba de editorial. Hoy en día, se pueden implementar estándares como JATS, y evitar los costes de I+D.

Por otra parte, se han ido desarrollado herramientas y aplicaciones en torno al estándar JATS, las cuales bajan los costos de tener un flujo de trabajo XML, por lo cual los costos hoy son mucho menores de lo que eran hace quince años. Esto permitiría que editoriales más modestas lograsen implantar XML y consecuentemente beneficiarse de todas las ventajas que XML puede proporcionar a sus revistas y otros contenidos.

El capítulo está dividido en cinco secciones, la primera está dedicada a introducir las diferentes formas de trabajo a nivel editorial cuando se implementan flujos que o bien se centran en JATS XML, o incorporan JATS XML en algún punto del proceso. La segunda sección se ocupa de editoriales de revistas en acceso abierto que implementan JATS y las diferentes modalidades de implementarlo. La tercera sección reúne ejemplos de editoriales académicas comerciales que optaron por JATS y se han convertido desde un estado anterior de formatos propios, o que directamente iniciaron su producción con JATS o con sus antecesores. La cuarta y última sección presenta un panorama sobre los usos actuales de JATS en las bibliotecas. Incluye ejemplos de bibliotecas nacionales, y también de bibliotecas académicas que ofrecen servicios editoriales. Los ejemplos que se presentan han sido escogidos para ilustrar la variedad de situaciones, pues se puede implementar JATS con o sin personalizaciones (ver capítulo 5), total o parcialmente, para producción editorial o para intercambio (o ambas cosas), etc.

En la última sección de este capítulo se desarrollan las principales características de los servicios editoriales en las bibliotecas, para luego en el capítulo siguiente, explorar si están utilizando XML para marcado en sus procesos editoriales, y en dicho caso, si implementan JATS o no. Esa información ha sido recabada a través de una encuesta vía correo electrónico.

### 6.1. Formas de trabajo con XML en los procesos editoriales

La forma de implementar XML refiere a cómo se integra XML en el proceso de producción editorial. Las grandes editoriales en la actualidad trabajan mayoritariamente a partir de ficheros en XML y esa versión está destinada a ser el formato de archivo definitivo de la publicación. Según Bazargan (2015), este requisito resulta en dos retos principales para las editoriales y para sus proveedores de servicios: crear un archivo XML a partir de lo que el autor ha enviado (ej.: un fichero Word o LaTeX), y generar a partir del contenido XML otros formatos: PDF, EPUB, etc. Agrega además que cada vez más la DTD de elección es JATS.

Los defensores del modo de trabajo centrado en XML (*"XML-centered"*), como Schwarzman y otros (2004), argumentan que el XML generado de esa forma es muy valioso: sirve como formato de archivo y además partiendo de éste como base, después que la calidad del XML está asegurada, los formatos secundarios como PDF, impresos o HTML se pueden derivar directamente de XML.

Mcglone (2013) opina que los flujos de trabajo XML tienen muchas ventajas, entre las que destaca el poder generar múltiples formatos de salida a partir del XML, utilizar los metadatos XML para catálogos, sitios web y bases de datos, automatizar procesos, poder acompañar el crecimiento de los servicios, así como también preservar el contenido digital para el futuro.

*XML workflows enable publishers to output content quickly and easily in several electronic formats (ePub, HTML, PDF); repurpose content into other channels (catalogs, websites, databases, printers); automate processes; scale their services and publications; and preserve the digital content for the future* (Mcglone, 2013).

Existen formas diferentes de abordar los flujos de trabajo con XML en los sistemas editoriales. En general, existen tres tipos de flujos de trabajo dependiendo del momento del proceso de publicación en el cual se comienza a trabajar con XML: al principio, en la mitad o al final (Publishers Press, 2016).

Integrar XML al principio (*XML First* o *XML-in*) es considerado el flujo de trabajo más eficiente, pero también puede ser el más difícil de lograr. Este tipo de flujo de trabajo XML comienza en la fase de autoría (Figura 12). Contar con el contenido marcado a medida que el autor lo va creando permite que el contenido sólo sea tratado una vez y todos los resultados (impresión, visualización en pantallas grandes o pequeñas, etc.) pueden ser creados a partir de este flujo único para los contenidos del documento. Los llamados *XML-First* son aquellos sistemas en los que no hay transformaciones previas, sino que parten de marcar el texto del documento desde que es creado.

XML en la mitad (*XML in the middle*) es la denominación que se utiliza cuando el marcado XML se realiza después que el proceso de edición se ha completado (Figura 13). Es considerado más fácil de implementar que *XML First* y también es visto como un paso intermedio, ya que una vez establecido este flujo, puede facilitar el camino hacia una solución *XML First*.

XML al final (*XML Last* o *XML-out*) es el flujo de trabajo más sencillo de implementar. El marcado tiene lugar después que la versión final del archivo está terminada, o se hace a partir de los archivos PDF (Figura 14). Esta forma de trabajo es la que adoptan editoriales que reciben los PDF u otros formatos de los artículos los cuales le son enviados por el editor de la revista, o también se aplica en conversiones retrospectivas de artículos de archivo de revistas que mantenían sus acervos en otros formatos, y cambian posteriormente sus flujos de trabajo para comenzar a utilizar XML.

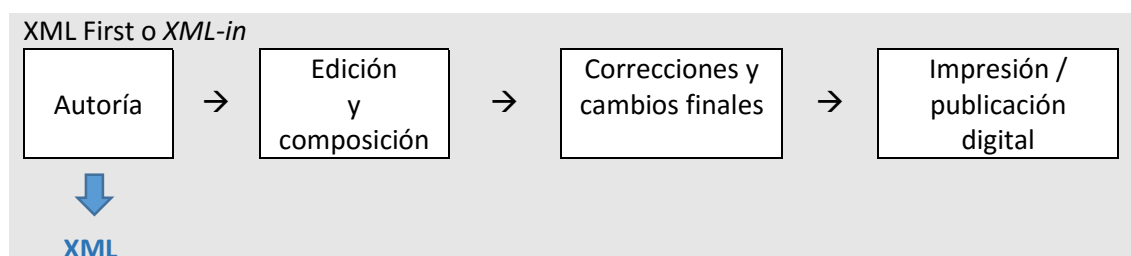


Figura 12. *XML First* o *XML-in*

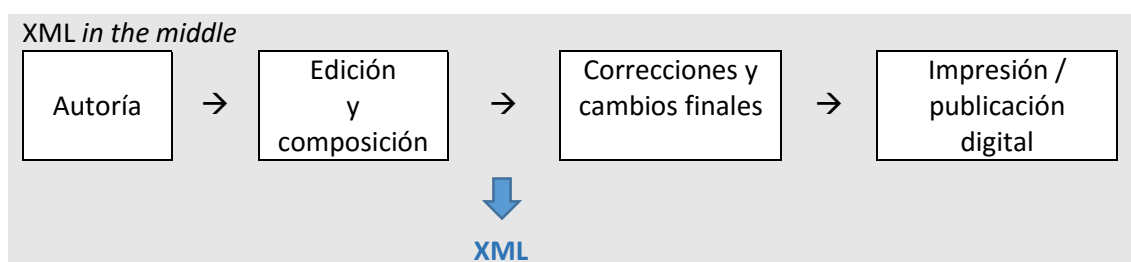


Figura 13. XML in the middle

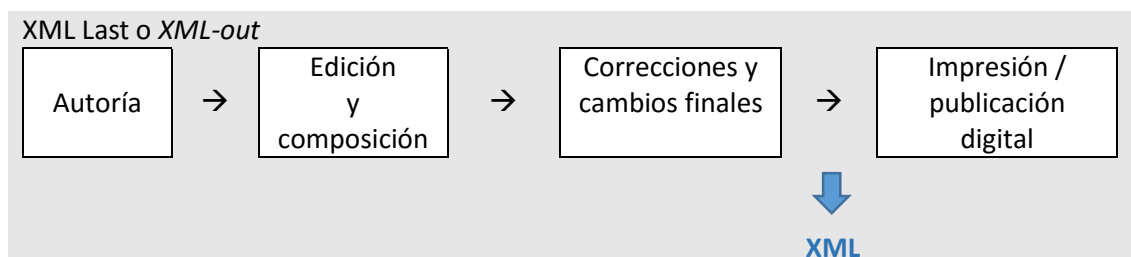


Figura 14. XML Last o XML-out

Figuras 12 – 14: Elaboración propia basada en Publishers Press (2016) <<http://www.pubpress.com/faq/how-can-publisher-integrate-xml-workflow>>

Si bien existen coincidencias sobre de qué se está hablando cuando se menciona un flujo *XML-first*, existen diferencias en cómo se aplican las tecnologías para implementar un flujo de estas características.

*We believe that XML-first should be interpreted in its more “pure” form, meaning that every PDF generated, including proofs going to authors, is generated fully automatically from the corresponding XML file, thus virtually guaranteeing one to one correspondence between the two files. We do admit this full automation is not an easy task and involves considerable programming, but it is possible and we believe the effort pays off long term* (Bazargan, 2015).

*Because XML is misunderstood, the method to institute an XML-first workflow is mired in confusion. XML is about defining and structuring content, and is applied through the normal editorial and production process. An XML-first workflow can be easily implemented using the typical software (e.g., Microsoft Word and InDesign) used within the publishing chain* (Rech, 2012).

Como menciona además el mismo Rech (2012), existen algunos problemas cuando se trabaja a partir de procesadores de texto, pues tanto *Word* como otros procesadores similares permiten la exportación directa de un documento a XML, pero esto no asegura que el XML resultante esté bien formado por más que se hayan aplicado correctamente los estilos en el procesador.

Respecto de las complejidades de los flujos de trabajo basados en XML, Maxwell y otros (2010) se preguntan si lo que sucede es que los flujos de trabajo basados en XML son simplemente demasiado complicados para ser adoptados fácilmente por las editoriales más pequeñas, y de ser así cuáles serán las implicaciones a medida que avanzamos en la era digital. La alternativa que estos autores proponen es un enfoque basado en flujos editoriales y de producción XML que utilice las herramientas ya existentes.

Dependiendo del tamaño del negocio y de las variables económicas del caso, la editorial puede realizar ella misma el marcado (marcado “*in-house*”) o encargarle esta labor a terceros (“*outsourcing*”). En opinión de Huh, Choi y Kim (2014), si se trata de una sociedad científica que publica más de cinco revistas o el número de artículos publicados por año es mayor a 1000, entonces es viable emplear a un especialista para producir archivos XML. El costo de producción de un artículo en XML aceptado internacionalmente según estos autores oscila entre 50 y 60 dólares americanos, aunque aclaran que puede variar dependiendo de la cantidad de tablas, formulas químicas y matemáticas que incluya el artículo.

Como expresan Schwarzman y otros (2004) debe tenerse en cuenta además al momento de contratar servicios externos que el marcado y la representación de contenido son funciones distintas, y no tienen por qué ser realizadas por el mismo proveedor de servicios. El marcado es llevado a cabo por un “*Tagger*”, quien produce los archivos XML a partir del formato que envía el autor. Otra es la función de quien produce los formatos de representación del contenido, como PDF y HTML a partir de XML. Además, ambas tareas pueden ser llevadas a cabo por la editorial o solo una de ellas.

Los editores se han visto limitados en sus opciones para establecer flujos de trabajo basados en XML por parte de sus sistemas de composición. Es una realidad que los sistemas de gama alta incluyen edición y composición del texto directamente en XML, pero son costosos y requieren de operadores capacitados. Por otra parte, los sistemas editoriales de escritorio que sí pueden ser capaces de aceptar entrada directamente en XML luego no tienen un buen desempeño gestionando los cambios realizados durante el ciclo de corrección, pues no logran que los mismos se reflejan fielmente en el XML. En opinión de O’Connor y otros (2015) el mayor desafío del proceso editorial basado en un fichero único -definida como la edición, composición y correcciones del texto, y publicación en varios formatos utilizando un único archivo como fuente- ha sido poder capturar los cambios que se realizan durante la fase de pruebas de página. Según estos autores, la clave para crear un flujo de trabajo basado en XML robusto y automatizado es poder realizar las pruebas directamente sobre XML, en un entorno de edición que requiera poco o ningún aprendizaje para el autor y proporcione un buen mecanismo de seguimiento y control de cambios al editor.

*The key to creating a robust, automated XML-based workflow is an XML proofing and editing environment that requires little to no learning for the author and has comprehensive change tracking for the editor. The solution to these limitations to single-source publishing is early creation of XML and use of a WYSIWYG XML editor coupled with rapid XML-to-PDF creation, with all components tied together through a production management system (O’Connor y otros, 2015).*

Es importante evaluar primero la conveniencia de la opción XML. No todo proyecto editorial requiere un flujo de trabajo XML o que sus documentos sean codificados en XML.

*Because the upfront costs to establishing an XML workflow can be quite considerable and not every digital publishing project requires XML encoded documents, defining the scope and aims of the project as early as possible can help minimize any unneeded efforts, and help to focus your publishing project's production processes (Gross, 2003).*

Serán el alcance y los objetivos del proyecto lo que determine finalmente si el proyecto requiere XML. Además, hay que tener en cuenta que los flujos XML se adaptan muy bien a la publicación de contenidos académicos tradicionales en varios formatos (basados mayormente en texto), a la reutilización en múltiples formas y preservación de contenidos a largo plazo. Pero a medida que la publicación académica en la web comienza incorporar varios tipos de contenido, interacción y

multimedia, XML y los flujos de trabajo basados en XML pueden no siempre ser la manera más eficiente de encarar formas de publicación académica no tradicionales (Mcglone, 2013).

## 6.2. JATS en editoriales y archivos con acceso abierto

La publicación de revistas se ha vuelto cada vez más diversa y potencialmente más competitiva con la aparición de nuevos modelos de negocio. En el modelo *Open Access* (OA) o acceso abierto, la investigación científica original es puesta a disposición con acceso libre en la red respecto a la mayoría de las restricciones sobre los derechos de autor y las licencias de reutilización (Ware y Mabe, 2015, p. 88).

El acceso abierto a los artículos tiene varias manifestaciones, que podemos agrupar en la vía dorada o "*Gold*": artículos en revistas de acceso abierto tanto si se pagó un cargo por la publicación de los artículos o no (incluyendo allí a las publicaciones completas en acceso abierto y a las híbridas: revistas de suscripción que dejan artículos disponibles en acceso abierto en cuanto se publican, por lo general como resultado del pago de un cargo por publicación), el acceso abierto con retraso: revistas de suscripción que dejan disponibles artículos en acceso abierto después de un período de embargo, y el auto-archivo ("*Green*" o vía verde): copias de las versiones de artículos, disponibles en repositorios institucionales o de otra índole, o en las páginas web de los propios autores. Pueden existir en múltiples versiones y estar archivados en diferentes repositorios.

Ware y Mabe (2015, p. 91 y 93) advierten sobre la diversidad de modalidades dentro de la vía dorada al acceso abierto, la cual se toma a veces como sinónimo del modelo de negocio de pago por publicación/*Article Publication Charge* (APC), pero estrictamente hablando el término vía dorada hace referencia a las revistas que ofrecen acceso abierto inmediatamente que el artículo ha sido publicado. O sea, la vía dorada incluye a los artículos de pago por publicación, pero además incluye otros modelos, por ejemplo, el apoyo institucional o el patrocinio. El modelo APC se ha vuelto más complicado, y con muchas variantes con acuerdos basados en la extensión, descuentos, anticipos y membresías institucionales, compensación y acuerdos para publicaciones híbridas, membresías individuales, etc. Dentro de estos modelos de negocio están comprendidas editoriales como PLOS ONE, PeerJ y eLife, todos los cuales implementan JATS.

La publicación en acceso abierto ha llevado a la aparición de nuevos tipos de revista, entre ellos el llamado *megajournal* o mega revista. Es un concepto nuevo del cual no existe una definición consensuada aún, pero sí que hay algunos autores que mencionan sus características. Según Momen (2014) la mega revista se caracteriza por tres rasgos fundamentales: el acceso abierto a texto completo con el cobro de una tarifa relativamente baja de publicación, la rápida revisión por pares basada en *soundness not significance* (juzgan los artículos por el rigor científico, no por la importancia de un artículo para la disciplina, su interés para la comunidad, la novedad, etc.) y por abarcar un ámbito temático muy amplio. El número de mega revistas continúa creciendo, y así queda demostrado en *STM Report 2015*, donde se listan más de 50 ejemplos (Ware y Mabe, 2015, p. 99, *Table 10*). Las grandes editoriales comerciales también han visto una mina en este modelo de negocio, con algunos ejemplos como Springer produciendo *Springer plus*, Sage publicando *Sage Open*, *BMJ Publishing Group* publicando *BMJ Open*. Pero según Momen (2014) ninguna de ellas podría ser considerada una mega revista dado que la cantidad de artículos es aún de unos pocos miles.

Los portales y proyectos OA que presentamos en esta sección responden a diferentes filosofías OA con sus diferentes modelos de negocio. Lo que tienen en común es que ofrecen al lector los artículos en acceso abierto y que implementan JATS en sus procesos editoriales. Están los que cobran



a los autores una tarifa por publicar, y los que no. Dentro de los que ofrecen sitio donde publicar sin que exista un cobro por prestar ese servicio se destacan los portales financiados por agencias nacionales, pidiendo como contrapartida que las revistas cumplan determinados requisitos de calidad, tanto de contenidos como de formato.

### PubMed Central

PubMed Central (PMC), es el archivo en acceso abierto de los artículos que contienen resultados de investigaciones financiadas por los Institutos Nacionales de Salud (NIH) de Estados Unidos. PMC retiene los derechos para archivar a perpetuidad todos los materiales depositados y se ha comprometido a mantener la integridad a largo plazo y la precisión de los contenidos.

La política de PMC está en concordancia con la política pública del gobierno de Estados Unidos respecto los mandatos sobre el acceso a los resultados de la investigación financiada con fondos públicos, en forma de publicaciones y sus datos subyacentes. Se publican en PMC documentos que deben estar disponibles por mandato legal<sup>11</sup>, pues los autores que publican resultados de investigaciones financiadas por NIH tienen la obligación de depositar sus artículos en PMC. Aquí es el propio sistema el que exige al editor de la revista ser capaz de producir archivos XML para enviar sus artículos a PMC. Aunque la revista no sea de libre acceso, el editor debe canalizar de un modo u otro esa demanda por una salida JATS XML, ya que es necesario para adaptarse a este requisito legal que debe cumplir el autor del artículo.

Si bien PubMed Central recibe artículos marcados con otras DTD, desde su sitio web recomiendan expresamente a los editores enviar el contenido de sus revistas en JATS XML, recomiendan *Publishing Tag Set* a los editores, y facilitan herramientas para validar, manuales y ejemplos de aplicación del estilo PMC. Si bien se recomienda que las contribuciones sean enviadas en XML conforme a *NISO JATS Journal Publishing Tag Set*, también aceptan XML conforme a otras DTD ampliamente utilizadas en la publicación de revistas científicas sobre ciencias de la vida. Cuando llegan artículos marcados con otros esquemas XML, los transforman a JATS antes de publicarlos, utilizando *Archival and Interchange Tag Set*.

En cuanto a los archivos a ser depositados, PMC requiere un archivo XML independiente para el texto completo de cada artículo, además de los correspondientes archivos de imágenes digitales originales, en alta resolución, para todas las figuras de cada artículo. Se puede incluir además un PDF del artículo si es que existe, pero nunca como la única forma. El envío puede incluir archivos de datos suplementarios (ej.: hojas de cálculo o archivos de vídeo) disponibles con el artículo.

### ScienceCentral

Otro ejemplo de implementación de JATS v. 1.0 “As-is” es ScienceCentral, un portal de acceso libre y gratuito financiado por el gobierno de Corea del Sur. Contiene el archivo a texto completo de revistas editadas por las sociedades científicas agrupadas en la Federación Coreana de Sociedades de Ciencia y Tecnología. Este portal implementa *JATS Journal Publishing Tag Set*, realizando un marcado completo con el estándar, y exigen a los editores de las revistas que envíen los contenidos en JATS XML.

ScienceCentral remite al sitio oficial de JATS para ampliar la información sobre como producir archivos en XML. La guía de marcado que aparece en el sitio ScienceCentral se basa en en la versión

---

<sup>11</sup> NIH Public Access Policy Details <https://publicaccess.nih.gov/policy.htm>

más reciente NISO JATS *Journal Publishing DTD*. Proveen un validador de estilo (*ScienceCentral Style Checker*) disponible en línea para que los editores verifiquen la calidad del marcado de sus ficheros JATS XML.

La forma de trabajo de ScienceCentral es comparable a PubMed Central en la mayoría de sus características y requisitos; sin embargo, hay dos diferencias. En primer lugar, su ámbito de publicación de revistas comprende todos los campos de la ciencia, y en segundo lugar, acepta revistas en todos los idiomas (Huh, 2013).

Para ScienceCentral la característica multi-idioma de JATS es uno de los factores que hace posible establecer un formato único de marcado independientemente de los idiomas presentes en cada artículo de la revista, ya que si bien la mayoría de las revistas del portal publican en inglés, otras lo hacen en coreano. Una de las características más notables introducidas por JATS versión 1.0 es la aceptación de todos los idiomas en el mundo. Los *Tag Sets* actuales del estándar JATS continúan proporcionando esas estructuras para la codificación de documentos en los que algunos de los metadatos o el texto estén en varios idiomas. Mediante la utilización del atributo “@xml: lang” dentro de las etiquetas se logra que contenidos en múltiples idiomas puedan coexistir en un mismo documento. Además, el uso del atributo “@id” dentro de los elementos permite la sincronización entre versiones traducidas. Sin embargo, Perera (2016) es de la opinión de que en general es más práctico utilizar el atributo “@id” para sincronizar el contenido a través de múltiples idiomas y múltiples ficheros que mantener mezclado el contenido en los diferentes idiomas en un solo fichero utilizando el atributo “@xml: lang”.

Según Huh y otros (2014), en Corea el obstáculo más común para la adopción de JATS es la falta de conocimiento acerca del estándar por parte de los editores. Según este autor, si un editor quiere producir un archivo JATS, es fácil y simple, pues el costo de la producción de los ficheros XML es considerado razonable.

Según las especificaciones que brinda el propio portal, para remitir artículos a ScienceCentral, se deben proporcionar los siguientes metadatos en cada fichero JATS XML: ISSN de la revista, editorial, declaración de derechos de autor, licencia de uso, volumen y número (si tiene), números o secuencia de páginas de cada artículo y la fecha de publicación. Deben enviarse además las figuras (TIFF o EPS, con al menos 600 dpi) y otros materiales complementarios. También se pide enviar un archivo PDF que corresponda con el fichero JATS XML, a los efectos de comprobar la exactitud del contenido de los archivos JATS XML. Los archivos PDF deben estar en formato digital legible.

## PLoS ONE

PLOS ONE (PONE) es una publicación internacional arbitrada, en acceso abierto. Forma parte de PLoS (*Public Library of Science*), una editorial sin fines de lucro con sede en San Francisco, California, Estados Unidos. PLoS es el buque insignia de la publicación de acceso abierto, publica en la actualidad siete revistas de acceso abierto arbitradas por pares, se caracteriza por estar disponible solo en formato digital, y se publica desde 2006. La cantidad de artículos revisados, aceptados y publicados se cuentan por miles. Es un *open access “mega-journal”* que implementa JATS, siendo una de las revistas que más artículos publican en la actualidad.

PLoS ONE es una revista de amplio alcance en las ciencias biológicas y médicas, tiene normas que son explícitamente diferentes de otras revistas científicas. Comparte estos criterios en su sitio, especificando que los materiales remitidos para publicación deben ser inéditos, basados en resultados de investigación, y “técnicamente sólidos y dignos de figurar en el registro científico”. Una vez que el

trabajo se publica en PLoS ONE, es la comunidad en general la que discute y evalúa su relevancia a través del número de citas que atrae, las descargas que logra, la cobertura que recibe en otros medios, y la post-publicación de notas, comentarios y puntuaciones que recibe en PLoS ONE (Sharp, 2013).

PLOS utiliza *NLM Journal Publishing DTD* versión 3.0 como su formato fuente en XML. PLoS ONE recibe contribuciones en Word, LaTeX, o RTF, que luego convierte a XML de acuerdo a *NLM Journal Publishing DTD* v3.0 antes de su publicación (Graham, 2014). El principal formato de publicación en las revistas PLoS es *XHTML 1.0 Transitional* con muchos enlaces de hipertexto, por ejemplo en la sección de referencias.

## PeerJ

*PeerJ* representa otro tipo de modelo de negocio dentro de las editoriales que implementan JATS. Publica revistas en acceso abierto y preprints en el área de biología, medicina e informática. Difiere de otros modelos de acceso abierto que cobran a los autores por publicar en que su sistema de cobro no es por artículo, sino en base a una membresía por autor, lo cual según sus creadores, resulta en una reducción de costos de publicación para los investigadores. *PeerJ* cobra a los autores una única vez, y les permite publicar en la revista el resto de su carrera. Dependiendo del nivel de la suscripción el autor puede remitir 1 o más artículos por año. Existen además requisitos adicionales, como que el autor debe comentar o revisar por lo menos un artículo de otros autores cada año. La baja en los costos de publicación la logran mediante el uso de la infraestructura de la nube, ejecutando sus servicios en Amazon EC2<sup>12</sup>, y almacenado el contenido en Amazon S3<sup>13</sup>.

En cuanto a la aceptación de los artículos para publicación, la investigación presentada se juzga basándose únicamente en la solidez científica y metodológica (como en PLoS ONE), y las revisiones por pares son publicadas junto con los artículos.

Respecto a los formatos de remisión de artículos, PeerJ acepta .DOC (MS Word), .DOCX (MS Word, preferido), .ODT (OpenDocument), .RTF (Rich Text Format, MS Word) y .PDF para usuarios de LaTeX. Dentro de las indicaciones a los autores se menciona que el texto del artículo debe ser formateado con la plantilla "Normal" (encabezados 1, 2, 3, etc., para las secciones). Se advierte que el uso de plantillas inusuales puede afectar la calidad del PDF de revisión. Solicitan a los autores no justificar el texto de su documento, sino alinear todo el texto a la izquierda. Todas esas indicaciones están orientadas a la compatibilidad con su sistema de conversión automática a JATS XML. Implementan *JATS Article Publishing DTD* y utilizan las herramientas de PMC para validar la calidad de los archivos JATS XML.

Otra característica de la forma de trabajo de PeerJ es que el autor debe llenar un formulario con metadatos descriptivos del trabajo que está cargando. Este requerimiento apunta a bajar los costos de producción, pues esos metadatos luego facilitan la creación del XML.

## eLife

eLife presenta otro modelo de negocio diferente en cuanto a publicación en acceso abierto. En este caso, la publicación es financiada por fondos que provienen de patrocinios a la investigación

---

<sup>12</sup> Amazon EC2: *Amazon Elastic Compute Cloud*. Es un servicio web que proporciona capacidad de cómputo de tamaño variable en la nube.

<sup>13</sup> Amazon S3: *Amazon Simple Storage Service*. Es un servicio web de almacenamiento de archivos en línea ofrecido por Amazon.

(*funders of research*). Se publican en esta plataforma artículos de investigación en ciencias de la vida y biomedicina.

Según Harrison (2016), cuando eLife fue lanzado en 2012, el sistema de envíos de contribuciones a las revistas se planificó para recoger los metadatos descriptivos a través del propio autor utilizando un formulario en las pantallas iniciales de envío del documento. De esa forma, el encargado de procesar el contenido no tiene que tomarse tiempo en buscar la información de las diferentes partes del manuscrito y otra información clave incluida en cualquier parte dentro del texto, por ejemplo los detalles de la financiación de la investigación. Los datos ingresados por el autor se convierten posteriormente a JATS XML cuando el artículo es aceptado para publicación. Se ha logrado agilizar el sistema de producción, bajando los tiempos y costos a través de la recolección de la mayor cantidad de metadatos posibles en el *Submission process system* montado en *Amazon web services* (Harrison, 2014).

El proceso editorial comienza por la recepción de artículos en PDF para revisión, y para simplificar el proceso de carga, piden a los autores subir su manuscrito completo como un único PDF. Además, conjuntamente con el PDF los autores deben enviar información adicional en la carta de presentación. Si bien eLife tiene por objeto alcanzar un flujo de trabajo *XML-first* para todos los metadatos, han encontrado más dificultades de las previstas para lograrlo, ya que los sistemas de envío no estaban preparados para recoger los metadatos con el nivel necesario para la producción y publicación.

Cuando el artículo es aceptado, el autor debe enviar el archivo de texto con las tablas principales al final, pudiendo optar entre RTF, DOC, DOCX, o LaTeX. Los datos se convierten a JATS XML al ingreso en el sistema de producción. El fichero de texto del autor pasa al proveedor de servicios de procesamiento para generar el XML del texto completo. Uno de los inconvenientes detectados en esta etapa, según Harrison (2016) es que algunas piezas de información no podían ser generadas directamente en JATS XML desde el fichero Word, específicamente algunos metadatos proporcionados por el autor en la carta de presentación. Como forma de solucionarlo, estos detalles se enviaban a través de “notas” al proveedor de servicios de procesamiento, lo cual requería la interpretación y la introducción de intervención manual. El otro inconveniente que pone de manifiesto Harrison (2016) respecto a este sistema, es que cuando un autor envía nuevamente un artículo y actualiza su fichero Word, no necesariamente actualiza la información en el formulario de envío, por lo que si los metadatos cambian tienen que ser actualizados por separado. Por lo tanto, debieron agregar una etapa de control de calidad antes de transformar el contenido a XML, a modo de asegurar que la versión final del fichero Word del autor se corresponda también con los datos proporcionados en el formulario y sea añadido como nota para quien se encarga de procesar el contenido. Eso resultaba costoso teniendo en cuenta la cantidad de metadatos que estaban recogiendo.

En 2016 eLife ha lanzado su propia plataforma de publicación. Con la nueva plataforma aceleran los tiempos de producción y añaden nuevas características al flujo de trabajo, como la incorporación de XML desde el ingreso del documento en el flujo de trabajo y la generación de versiones PDF del documento en cualquier momento, de forma automática en cada punto del proceso de producción. Mediante la nueva plataforma, todas las acciones se realizan en línea, tanto si se trata de producir, procesar contenido, editar texto o si el autor necesita realizar una modificación.

Según Harrison (2016) actualmente se encuentran en proceso de revisión de los datos solicitados en las pantallas de envío, con el objetivo de reducir la cantidad de información solicitada actualmente a los autores. Sin embargo, como el flujo de trabajo para la publicación de los manuscritos

todavía se basa en estos datos, no les ha sido posible aún eliminar del todo este requerimiento para pasar a basarse completamente en el fichero remitido por el autor.

## SciELO

SciELO es un proyecto que comenzó en 1998 con la cooperación entre la *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP), la Biblioteca Regional de Medicina (BIREME) y los editores científicos de Brasil. Actualmente es un modelo de publicación de revistas científicas en formato electrónico no solo en Brasil sino también en Iberoamérica y Sudáfrica. Las colecciones SciELO son multidisciplinarias y multilingües.

En este portal, la mayoría de las revistas son gestionadas de forma independiente por las sociedades científicas o instituciones académicas que las editan. FAPESP proporciona el apoyo político y financiero para el desarrollo de la colección SciELO Brasil, que interactúa con las otras colecciones nacionales y temáticas. SciELO Brasil proporciona el mantenimiento de la plataforma tecnológica y es quien coordina y actúa como secretaría técnica de la red. Cada una de las colecciones nacionales SciELO es administrada y financiada por su correspondiente institución nacional de apoyo a la investigación. Todas las colecciones siguen la misma metodología y utilizan la misma tecnología (Packer y otros, 2014).

La función editorial de SciELO consiste en proporcionar una plataforma en acceso abierto en la web, si bien como mencionan Packer y otros (2014) ha habido además una expansión progresiva de las funciones editoriales, incluida la prestación del servicio de gestión de envío de los documentos en línea, la incorporación de XML a texto completo como fuente de HTML, PDF y ePUB como formatos de publicación, y la asistencia editorial y en la difusión para los editores de revistas.

SciELO es uno de los ejemplos de implantación de personalizaciones de JATS en la actualidad. Desde 2015 implementan SciELO PS y no reciben artículos que no estén marcados de acuerdo a su esquema.

*A plataforma SciELO deixará de aceitar arquivos em outros formatos como, por exemplo, Word, PageMaker, InDesign. O padrão de formatação de texto do SciELO baseia-se no JATS (Journal Article Tag Suite) do PubMed Central (PMC) e envolve a marcação de tags ao longo do arquivo para que informações relevantes dos artigos sejam automaticamente identificáveis* (Kimura y de Araujo Machuca, 2014).

En opinión de Kimura y de Araujo Machuca (2014), ambos editores de una revista de ciencias sociales aplicadas que publica sus contenidos en SciELO, si bien el cambio vigente desde 2015 (momento en el cual los editores debieron comenzar a proporcionar los ficheros en formato JATS XML) es sustancial, las ventajas que proporciona dicho formato hacen que valga la pena asumirlo. Dentro de las ventajas que nombran están: 1) Las métricas de productividad se pueden generar fácilmente, proporcionando los indicadores pertinentes para evaluar no sólo el rendimiento sino también la producción de los investigadores, la influencia de las instituciones y el impacto de los proyectos de investigación; 2) Los cálculos de impacto de revistas, como el Índice h de autores, se pueden realizar de forma automática a partir de los artículos con marcado XML; 3) Además, el marcado XML también permite adaptar fácilmente la presentación del trabajo a diversos dispositivos, como ordenadores, tabletas y teléfonos inteligentes; 4) La difusión del conocimiento, pues el marcado XML implica mayores posibilidades de acceso a los artículos, dando mayor visibilidad a la labor de los investigadores.



En el caso de los artículos a publicar en SciELO, la validación de los ficheros XML se realiza en local con herramientas que se instalan en el ordenador y luego de validado el artículo es enviado para su publicación vía FTP. Mediante SciELO Manager<sup>14</sup>, el sistema de gestión de títulos, números y secciones de revistas (se accede a él en línea mediante usuario y contraseña) se crean el número de revista y sus secciones (también el título, si se tratase de una nueva revista), permitiendo luego de completado el registro, descargar un .zip con los metadatos correspondientes para el marcado de las contribuciones a ese número. Esos metadatos, previa extracción del .zip y copia de los ficheros en el directorio local correspondiente, son tomados por *Markup*, una macro de Word instalada de forma local en el ordenador del usuario. Tanto las herramientas para validación y conversión como la macro de Word para hacer más sencilla la marcación de los artículos, son parte del paquete de software que SciELO pone a disposición de los editores de las revistas que envían artículos para su publicación en el portal. Las aplicaciones informáticas anteriormente mencionadas fueron desarrolladas con el objeto de asistir con la generación y la validación de los ficheros a personas que no tienen conocimientos de XML. Todas ellas son proporcionadas y mantenidos por SciELO, siguiendo el modelo de código abierto.

*Markup*<sup>15</sup> es una de esas herramientas. Como acabamos de decir, es una aplicación de escritorio (macro en *Microsoft Office Word*). Dentro de sus funcionalidades se encuentran: identificar los elementos de artículos y textos en un documento de Word, identificar semiautomáticamente los elementos de las referencias, generar y también validar archivos XML según SPS y JATS DTD.

Además de *Markup* existen estas otras aplicaciones como *Stylechecker*, una herramienta web que ofrece informes detallados sobre el cumplimiento por parte del fichero XML objeto del chequeo de la especificación SciELO PS. Otro ejemplo de herramientas para facilitar el manejo de ficheros XML a los editores son *Packtools* y *Package Maker*, ambas aplicaciones para correr en local. *Packtools* es una biblioteca de *Python* que añade características y utilidades para manipular paquetes SciELO PS y ficheros SciELO PS XML. *Page Maker* es una herramienta para la generación de paquetes SciELO PS y PMC (sólo funciona en Windows). Genera informes detallados sobre la estructura y la validez de algunos de los metadatos de los documentos XML.

## J-STAGE

J-STAGE es un ejemplo de archivo en acceso abierto que implanta JATS “tal cual es”, pero parcialmente: solamente marca metadatos de la cabecera y de la bibliografía del artículo. Las revistas de J-STAGE son administradas por cada sociedad científica, las cuales tienen su propio proceso de envío de artículos y sus normas de presentación. La mayoría de las revistas envían sus PDF al portal, pero algunas de ellas implementan otra modalidad basada en J-STAGE *Online Submission and Review System*, una aplicación en línea brindada a través de la aplicación *Editorial Manager* (utilizan los servicios de la empresa *Scholar One Manuscripts*).

J-STAGE es una de las principales plataformas de publicación de revistas electrónicas de Japón, iniciada en mayo de 2012 por *Japan Science and Technology Agency* (JST). J-STAGE ha adoptado JATS y requiere que cada revista incluya los metadatos y la información bibliográfica en formato JATS (Keishi, Tokushige y Mitsuru, 2012). Para acompañar este enfoque editorial han desarrollado una herramienta de conversión automatizada de PDF a JATS y se embarcan así en lo que Keishi y otros (2012) denominan la “XML-ization” de las revistas académicas en Japón.

---

<sup>14</sup>SciELO Manager: <http://manager.scielo.org/accounts/login/>

<sup>15</sup> Markup plug-in: <http://docs.scielo.org/projects/scielo-pc-programs/en/latest/download.html>



Al ser JATS XML un requisito, el portal ofrece una aplicación informática a los editores de las revistas de las sociedades científicas que envían artículos a publicar en la plataforma J-STAGE de JST. Esta aplicación se encarga del análisis automático de la versión final en PDF del artículo. La aplicación que ofrece J-STAGE es un servicio web que ayuda en la creación de metadatos JATS pues es capaz de extraer información bibliográfica directamente de los artículos y crear un fichero JATS XML con esos datos. Keishi y otros (2012) aseguran que la conversión automatizada está alcanzando una precisión del 90% y los planes futuros son expandirla y poder producir texto completo en XML a partir de los PDF.

Dentro de las limitaciones que Keishi y otros (2012) mencionan para esta aplicación está que no tiene capacidad para hacer frente a páginas escaneadas o PDF que no hayan sido generados con OCR, e incluso si han sido generados así, si la tabla de códigos de caracteres no está embebida en el PDF, el programa no tendrá la información necesaria para realizar la conversión. Dentro de las facilidades que brinda la aplicación está la función de edición de XML, con la cual el contenido XML transformado de forma automática se puede comprobar y editar. De esta forma, si el contenido ha sido incorrectamente reconocido por la aplicación, se puede corregir. Del mismo modo, si hay elementos para los cuales la aplicación ha tenido dificultades con la extracción automática, es posible introducir los elementos de forma manual en la ventana de edición.

### 6.3. JATS en editoriales académicas comerciales

Las editoriales agrupadas en esta sección para el estudio de la implantación de JATS, son editoriales cuyo modelo de negocio principal es el pago por suscripción, o sea publican revistas basadas en el modelo de suscripción y facilitan acceso a la versión digital de las mismas con o sin período de embargo, aunque esto no excluye que tengan sistemas híbridos (que permiten OA) y también algunas opciones de acceso libre a determinados materiales. De las editoriales académicas que seleccionamos, algunas tienen un volumen de negocio muy grande y varias de ellas son referidas como *mega-publishers*.

Reed-Elsevier, Wiley-Blackwell, Springer y Taylor & Francis han aumentado su participación en el mercado editorial, especialmente desde el advenimiento de la era digital. Las cuatro editoriales anteriormente mencionadas son los líderes del mercado editorial académico, y el quinto lugar en el *top five* lo comparten dos editoriales de diferentes áreas: Sage en Ciencias Sociales y Humanidades y *American Chemical Society* (ACS) en Ciencias Naturales y Medicina. En conjunto, estas cinco editoriales son las más prolíficas. Entre las cinco publicaron más del 50% de todos los artículos en 2013 (Larivière, Haustein y Mongeon, 2015).

*On the whole, our results show that the top commercial publishers have benefited from the digital era, as it led to a dramatic increase in the share of scientific literature they published. It has also led to a greater dependence by the scientific community on these publishers* (Larivière y otros, 2015).

Grandes editoriales que ya venían trabajando con marcado estructural y semántico han migrado formato varias veces. Sólo en las últimas dos décadas, se ha utilizado SGML primero y luego XML, implantando variados esquemas, entre ellos NLM DTD y ahora JATS. Volcar el contenido de un formato a otro no es un ejercicio trivial, y no es barato. Además de dinero, se requiere una buena planificación y gestión para hacerlo bien. Según Lapeyre (2009) todos los usuarios de NLM DTD en 2007 utilizaban las DTD y no el esquema XML de W3C. Los usuarios a los que ella hace referencia son:

Portico, Wiley, Elsevier, ACS, AIP, IEEE, Springer, ACM, BioOne, *PubMed Central*, *British National Library*, *Library of Congress*, *Highwire Press Library of the Sciences and Medicine*, editoriales universitarias (*Oxford*, *Cambridge*, *Johns Hopkins*, *Chicago*), *Commonwealth Scientific and Industrial Research Organization* (CSIRO), además de editoriales de sociedades científicas que envían artículos a alguno de los archivos o agregadores anteriormente mencionados.

El panorama actual de las grandes editoriales comerciales respecto a la producción de revistas es apostar a la automatización de la producción editorial. Muchos de ellos se han convertido desde otros formatos digitales a XML, o han pasado de una forma de publicación tradicional a implantar un sistema de producción basado en XML. Es el caso de *SAE International*<sup>16</sup>, la editorial de la Sociedad de Ingenieros Automotrices (*Society of Automotive Engineers*), que agrupa a los ingenieros y las industrias del sector aeroespacial y de automoción dedicadas a la fabricación de vehículos. Ellos convirtieron masivamente una biblioteca entera de documentos de ingeniería al formato XML. No publicaban en formato digital con anterioridad a 2008 pero actualmente utilizan un proceso altamente automatizado que busca sacar el mayor provecho de la utilización de NLM XML, XSLT y *Adobe In Design*, resultando todo ello en mayor productividad y mejor calidad de los productos finales (Fadik, Gross y Trombley, 2015). Estas son editoriales que asumen los costos de migrar sus contenidos, pueden permitirse invertir en tecnología que les permite bajar los costos de producción, y apuestan a los beneficios que la inversión les devolverá a medio plazo.

Actualmente, empresas tecnológicas proveedoras de servicios editoriales ofrecen sus plataformas basadas en la web que se encargan de la gestión del proceso editorial completo, desde la autoría a la publicación, donde el archivo que se está trabajando es siempre JATS XML. Este método minimiza el riesgo de errores en la conversión de formato y además acelera el proceso de publicación (Bazargan, 2015).

El flujo de trabajo de estas plataformas alojadas en la web se basa en un solo archivo JATS en la nube, guardado conjuntamente con los gráficos y material complementario. Utilizando un control basado en roles, diferentes grupos de interés se registran en el sistema y de esta forma realizan las modificaciones en el XML, utilizando una interfaz fácil de usar que se ejecuta en el navegador.

### **Nature Publishing Group/Palgrave Macmillan**

Otro ejemplo del avance de la implantación de JATS es la adopción del estándar por el grupo editorial *Nature Publishing Group/Palgrave Macmillan*. Este grupo comenzó a evaluar la migración en 2009, y como resultado ha implantado en 2012 un programa de producción y migración gradual retrospectiva de sus fondos a JATS que incluye: introducir JATS XML en el flujo de trabajo de todos los contenidos nuevos y crear una nueva plataforma de publicación que muestre los artículos JATS XML directamente en su plataforma, reduciendo con esto los tiempos de publicación de los nuevos artículos.

*A significant aspect of this program was to develop a mature journal article data model, which catered for all our data structures, whilst being flexible, extensible and comprehensible. We have adopted the JATS Journal Publishing DTD as it fulfils all of these requirements* (Donohoe, Sherman y Mistry, 2015).

---

<sup>16</sup> *SAE International*: <http://www.sae.org/>

Macmillan está invirtiendo fuertemente en la construcción de una plataforma de publicación escalable moderna y JATS es el modelo de contenido XML escogido. Han comenzado además la enorme tarea de convertir el contenido de archivo a JATS.

La publicación de revistas en Macmillan Publishers en 2008 era llevada a cabo por dos editoriales: Nature Publishing Group y Palgrave Macmillan. Donohoe y otros comentan que en ese entonces trabajaban con cinco proveedores de XML, y tres DTD propietarias poco documentadas y mal diseñadas. Con el objetivo de mejorar la eficiencia del flujo de trabajo, y ampliar el número de revistas que publicaban, se hacía necesario reconstruir la mayoría, si no todas las herramientas del sistema editorial a partir de cero, y eso se simplificaba si había un solo esquema XML. “Queríamos una DTD existente que pudiéramos adoptar y adaptar a nuestras necesidades con el mínimo esfuerzo. NLM DTD ya estaba siendo utilizada por otras editoriales, y estaba bajo desarrollo activo, por lo que parecía una elección obvia.” (Donohoe y otros, 2015).

El mapeo de las DTD propietarias a JATS se completó a principios de 2012, con el objetivo de comenzar a utilizar JATS para los contenidos anteriores, además de introducir un flujo de trabajo JATS XML para todo el contenido nuevo. Mostrar el contenido anterior todavía no convertido a JATS (válido para dos de las tres DTD propietarias: AJ DTD y NPG DTD) en la nueva plataforma les significó crear conjuntos de transformaciones XSLT para convertir NPG a JATS y AJ a JATS, y construir las herramientas para realizar ese proceso.

*We chose the blue Journal Publishing DTD because it offered just the right level of structure. One of the prime aims of migration to a new DTD was to simplify the code required for processing the XML, and we felt that the Archiving version was too loose, and would have required extra Schematron rules to enforce correct markup* (Donohoe y otros, 2015).

El flujo de trabajo actual es un flujo híbrido para alimentar el nuevo portal, que contempla la producción y publicación de nuevos contenidos en JATS, y los procesos que se necesitan para mostrar los contenidos marcados con las DTD anteriores en el nuevo portal, mientras se continúa el proceso de conversión de todo lo publicado con las DTD propietarias a JATS. En suma, están convirtiendo su contenido antiguo a JATS para unificar, desarrollaron las herramientas de conversión, y además generan el contenido nuevo directamente en JATS XML para el nuevo portal web de la editorial. Para esto último, utilizan JATS XML “tal cual es” y validan reglas propias con *Schematron*.

## De Gruyter

El grupo De Gruyter es una editorial académica con sede en Alemania, de 260 años de trayectoria, que publica libros y revistas científicas en las áreas de humanidades, ciencias sociales, ciencia, tecnología y medicina (STM) y legislación.

De Gruyter adoptó el esquema JATS / BITS para sus contenidos y estableció sus directrices XML específicas para la creación de los metadatos XML y los datos a texto completo. Ha desarrollado una aplicación para validar la calidad de los datos de los paquetes de ficheros de libros y revistas que le entregan sus proveedores de servicios. La herramienta de validación, que se basa en software de código abierto Talend<sup>17</sup> y Transpect<sup>18</sup>, verifica la consistencia de los metadatos, valida contra el

---

<sup>17</sup> *Talend Open Studio for Data Integration*: es una aplicación de código abierto desarrollada por Talend y diseñada para combinar, convertir y actualizar datos. Disponible en: <http://sourceforge.net/projects/talend-studio/>

<sup>18</sup> *Transpect*: aplicación de código abierto para la conversión y la comprobación de datos. Utiliza tecnologías estándares como XSLT 2.0 y XProc. <http://transpect.github.io/>

esquema JATS y reglas de De Gruyter, las cuales se especifican con *Schematron*. La aplicación devuelve un informe en HTML que proporciona información sobre los archivos de origen y sus respectivos mensajes de error. Los mensajes contienen la ubicación de error y se agrupan por su gravedad. Los contenidos que pasan el chequeo son enviados para su archivo y publicación. Este proceso garantiza una representación correcta de los contenidos en el sitio web de la editorial y facilita su recuperación y tratamiento en el futuro (Kraetke y Buehring, 2015).

### American Chemical Society Publications Division

La *American Chemical Society Publications Division (ACS Pubs)*, publica cerca de 50 revistas de la Sociedad Americana de Química, ofreciendo a la comunidad científica de todo el mundo la investigación de vanguardia publicada en una de las revistas revisadas por pares más citadas en química y ciencias afines.

Una vez que el manuscrito ha sido revisado y aceptado para su publicación, *ACS Pubs* recibe el formato de documento presentado por el autor (ej.: Microsoft Word o LaTeX). El formato interno para los manuscritos es XML, para facilitar los procesos de producción y publicación. Por lo tanto, antes de que ocurra cualquier proceso, el documento original del autor se envía a un proveedor para su conversión a XML, usando una DTD que es una versión personalizada de NLM DTD (Stoker y Rose, 2012).

O'Brien, Fisher y Haines (2011), explican que se realizó una personalización de NLM *Tag set* al nivel *Built from* implementado mediante modificaciones, la cual resultó en el perfil personalizado denominado *ACS Journal Tag Set*.

*While the NLM tag set served as the starting point for development of the ACS Journal Tag Set, the original tag set modules themselves were modified to implement the ACS-specific requirements. The ACS Journal Tag Set and DTD are represented in the customization profile with a customization level of "Built from" and customization implementation method using "Modifications" (O'Brien y otros, 2011).*

Oakes (2014) reporta que el sistema automatizado para la composición de artículos de revistas, el cual se puso en marcha en 2012, actualmente se encarga de la producción para todo el programa de revistas de ACS, gestionando más de 40 títulos y 300.000 páginas por año. El flujo de trabajo es *XML-first* (JATS / NLM) y todos los manuscritos están en formato XML desde la aceptación. Con ese sistema, las pruebas y las versiones finales se componen sin interacción humana para más del 90% de los artículos, lo cual reduce los costes y los tiempos de producción.

### Cambridge University Press

Cambridge University Press, es la segunda editorial académica más antigua del planeta, con un volumen de publicación importante tanto en libros como revistas. Es el caso que hemos tomado como ejemplo de implantación de BITS "tal cual es". En su implantación se tuvo en cuenta, entre otros aspectos el control de reglas propias con *Schematron*, el uso de MathML y otros estándares similares.

Según Eden y Cleghorn (2016), la utilización de marcado para los contenidos de libros académicos en *Cambridge University Press* a lo largo de su historia dio lugar a un modelo propio, basado en NLM DTD. A medida que aparecen nuevos requerimientos y nuevos tipos de contenido, éstos debieron irse añadiendo a la DTD y a las reglas de validación relacionadas. Pero cuando los cambios se vuelven frecuentes e impredecibles, crean puntos débiles en los flujos de producción, tanto

a nivel interno para mantener los procesos automatizados actualizados, como a nivel externo para mantener actualizados los conocimientos de los que les proveen servicios editoriales.

Luego de revisar la situación, se llega a la elección de BITS como la DTD a utilizar, en lugar de definir DTD propietarias. La decisión se basa en que BITS es un estándar de facto emergente de la industria, que está estrechamente alineado con el estándar NISO para revistas (NISO JATS Z39.96-2012), ya en uso en *Cambridge University Press*. También se considera que el estándar se ha afianzado, siendo de esperar ciertos beneficios futuros derivados del aporte y los requerimientos de otras editoriales (Eden y Cleghorn, 2016).

En este caso, la editorial ha modificado los flujos de trabajo editorial para implantar BITS, y se ha puesto énfasis en el control del marcado en todas las etapas del proceso. Según Eden y Cleghorn (2016), esos cambios dieron lugar a un proceso de trabajo editorial considerablemente mejor definido y previsible para los proveedores externos y también para el flujo interno dentro del sistema de la editorial, dando lugar a la creación de una base sólida para la producción de libros en el futuro.

## 6.4. JATS en bibliotecas

En esta sección se presenta un panorama actual de la utilización de JATS por parte de las bibliotecas. Destacamos aquí varios ejemplos de implantación agrupándolos dos apartados. El primero reúne ejemplos de bibliotecas que utilizan JATS fundamentalmente para archivar, preservar y para brindar servicios de acceso a contenidos. El segundo apartado se centra en el uso de JATS para brindar servicios editoriales.

*The recent burst of interest from the library community is largely due to proposals to use XML in e-journal archiving and also a growing awareness of the advantages of XML metadata. But the publishing community are[sic] increasingly viewing XML as playing a central role in the future of the entire e-journals process (Wusteman, 2003a).*

### 6.4.1. Archivo, preservación y acceso

En 2006 la Biblioteca del Congreso de Estados Unidos y la Biblioteca Nacional Británica comienzan a apoyar abiertamente la adopción de un estándar común, bajo el entendimiento de que NLM iba a formalizar en un futuro cercano su *NLM Tag Suite* y los modelos para contenidos de revistas con NISO (Library of Congress, 2006).

La preservación de los contenidos de las revistas es un objetivo común para bibliotecas y editoriales. La preservación de los trabajos científicos es fundamental para garantizar el acceso futuro. Tanto académicos, investigadores, bibliotecarios y editores reconocen la necesidad de que los archivos digitales se protejan y conserven para las futuras generaciones.

La preocupación de académicos y bibliotecarios sobre la disponibilidad de las revistas a largo plazo, y la tendencia de las bibliotecas a que sus suscripciones a revistas científicas estén cada día más orientadas al formato electrónico, hace más relevante la existencia de los archivos digitales de contenidos de revistas científicas (*e-journal archiving*), con vistas a su preservación digital. Su rol es preservar los ficheros y la presentación original de la editorial y encargarse de proporcionar los contenidos con presentaciones adecuadas a la tecnología actual.

La variedad de formatos de los objetos digitales en un archivo afecta el costo y la complejidad de las futuras migraciones de formato. Con el fin de controlar esto, una de las estrategias de preservación de los archivos es normalizar, utilizando estándares para los objetos depositados como



forma de reducir la variedad de formatos a un conjunto de formatos preferidos, siempre que sea posible. En cuanto a los formatos para los documentos, muchas editoriales de revistas electrónicas codifican el contenido del artículo en XML (o planean hacerlo pronto). Como ya hemos mencionado anteriormente, las editoriales utilizan diferentes esquemas, ya sean estándar (con o sin modificaciones) o propias, para satisfacer sus necesidades específicas. Desde el punto de vista de un archivo digital de revistas, escoger un esquema de marcado que permita normalizar todos estos documentos, implica reducir la complejidad de la documentación, la migración, y simplifica las cosas al momento de implementar una interfaz para acceso a los contenidos.

Un enfoque basado en estándares es importante para la preservación, pero también hay factores que inhiben su uso como estrategia de preservación digital. Como el ritmo de cambio es tan rápido cuando los estándares han alcanzado la etapa de ser aprobados formalmente – un proceso que suele tardar años– pueden quedar inevitablemente rezagados frente a otros desarrollos e incluso pueden ser superados. El desarrollo de extensiones propietarias o incluso las implementaciones de los estándares pueden diluir las ventajas de éstos en cuanto a la consistencia e interoperabilidad, lo cual afecta la preservación digital (Digital Preservation Coalition, 2015).

## Portico

Portico es un prestador de servicios de preservación digital especializado en contenido académico publicado en formato electrónico y ofrece tres servicios de preservación específicos, uno orientado a las revistas electrónicas, otro para los libros electrónicos y un tercero especializado en colecciones históricas digitalizadas (Beagrie, 2013). El servicio se financia con las cuotas anuales que pagan las editoriales y bibliotecas asociadas, garantizándoles que el contenido del archivo puede ser accedido en caso de que el acceso por los canales habituales se vea eventualmente interrumpido.

El esquema utilizado para normalizar los contenidos de revista es una versión propia basada en *JATS Archival and Interchange Tag Set* desarrollada por Portico. Para cada artículo de revista en el archivo, Portico mantiene todos los objetos digitales originales proporcionados por la editorial, incluyendo imágenes de páginas PDF, conjuntamente con los metadatos estructurales, técnicos, descriptivos y de los eventos (*events metadata*), además de una versión normalizada con el perfil de Portico basado en el conjunto de etiquetas de Archivo e Intercambio de NLM [actualmente JATS] del fichero del artículo proporcionado por el editor en SGML o XML (Morrissey et al., 2010).

Cada objeto archivado (un artículo de revista, un libro electrónico) puede ser empaquetado en un fichero ZIP, conteniendo todos los objetos digitales originales proporcionados por la editorial, junto con los objetos digitales creados por Portico y los metadatos XML asociados al objeto. El archivo completo puede ser reconstituido a partir de estos ficheros ZIP, siendo posible dar acceso a él a través de lectores y plataformas no específicos del proveedor, y completamente independientes del sistema de archivo de Portico (Morrissey, 2011).

*In general, the problems that Portico experiences with transforming JATS are really about transforming XML and not specific to JATS. When we received content in JATS from publishers we do transform it to our internal version of JATS [...] While JATS isn't a panacea on the transformation and migration front (we do still encounter publisher specific issues), it makes things much simpler than the old style of everyone doing their completely separate thing (Wheatley y otros, 2015).*



## Biblioteca Británica

El programa de depósito legal de revistas en formato electrónico recibe ficheros en diferentes formatos. Claire Caulfield, *Digital/Print Operations Manager* de la Biblioteca Británica, ha respondido al ser consultada vía correo-e acerca de los formatos utilizados por la Biblioteca: *I can confirm that the British Library uses the following DTDs when processing eJournal content: JATS, NLM Journal Publishing, Journal Archiving and Interchange, PubMed, PubAbs, MetaPress and different flavours of the above (such as versions and Atypen). We do also use others that are publisher specific* (C. Caulfield, comunicación personal, 12 de abril de 2016).

Además de conservar los metadatos del editor, los transforman para ser reutilizados en el Sistema interno y en los catálogos: *We keep the original metadata file as supplied by the publisher, but also transform it for use in our internal systems and to form part of the final catalogue record for the item.* (C. Caulfield, comunicación personal, 18 de abril de 2016).

La Biblioteca Británica (*British Library*) trabaja en colaboración con un proveedor externo de servicios de archivo para revistas electrónicas que ya hemos mencionado anteriormente en este capítulo: Portico. *Since 2013, the British Library has been working in partnership with Portico to provide e-journal content (including content in JATS format) for the Library's legal deposit programme* (Wittenberg y otros 2014).

Un informe publicado por el Grupo de trabajo de Preservación Digital de la Biblioteca Británica (*British Library Preservation Team*) sobre la evaluación de los riesgos de preservación para los contenidos almacenados en JATS XML, llega a la conclusión de que los mayores riesgos de preservación para los ficheros JATS son: la variabilidad y aspectos relacionados con la calidad (Wheatley y otros, 2015).

Según Wheatley y otros (2015), los riesgos relacionados con la calidad pueden mitigarse con una comprobación a fondo del contenido de los ficheros JATS XML y de su contenido asociado al momento de la recepción, mediante la implementación de un flujo de trabajo normalizado, que sirva también como instancia para la verificación minuciosa calidad del contenido depositado. Esto último aparece como esencial para asegurar la integridad del depósito, así como también para comprender la variabilidad en el contenido depositado por diferentes editoriales. La normalización oportuna a una única forma de marcado con JATS es el método utilizado para conseguir esa verificación de la calidad, haciendo frente al mismo tiempo a la variabilidad que puede presentarse en el contenido y con ello se evitan problemas posteriores de preservación. Además, estos autores recomiendan mantener una copia del paquete original depositado por la editorial, lo cual proporcionará algunas garantías respecto a eventuales problemas que se puedan presentar en la conversión.

El desafío más grande de preservación se presenta para los contenidos externos referenciados o insertados en los artículos. Es una realidad que este tipo de contenido externo está presente y continuará creciendo en la medida que los artículos científicos integran cada vez más otros productos de investigación, incluyendo los conjuntos de datos de las investigaciones. De no ser por esto, el carácter abierto y la buena interoperabilidad de JATS hacen que éste sea una buena opción para la preservación de artículos de revistas electrónicas (Wheatley et al., 2015).

## Biblioteca del Congreso de Estados Unidos

La *Library of Congress* (LC), Biblioteca Nacional y sede de la Oficina de Derecho de Autor de Estados Unidos, está muy involucrada la adquisición y gestión digital. En los últimos años, la Biblioteca del Congreso ha ido adquiriendo publicaciones seriadas electrónicas bajo los términos de la Ley de

Derecho de Autor y las ha ido agregando a su colección. Para conseguir este objetivo necesitan, además de adquirirlas, que estén disponibles en la plataforma de la biblioteca. El problema principal es que las publicaciones les llegan de las editoriales en muchos formatos, los cuales hay que normalizar, y además proteger las obras con *copyright* del uso indebido.

En abril de 2011, como parte del Programa *eDeposit*, la Biblioteca estableció formalmente el Proyecto *eSerials*. El objetivo del proyecto es asegurar que las publicaciones seriadas electrónicas publicadas en Estados Unidos y disponibles solamente en línea pasen a integrar las colecciones permanentes de la Biblioteca del Congreso a través del depósito obligatorio.

*The project team has met several times with publishers participating in the first phase of this effort. Based on their input in these conversations, we believe that the publishing community is moving to support JATS across the many subject areas in which the Library collects materials. We also believe the article-focused nature of JATS will become even more useful as more publishers move to article-based publishing. [...] Our goal is to acquire 80% of e-deposited content in JATS or other standard formats, and to normalize the other 20% of content that can't be acquired that way internally (Delfino y Mandelbaum, 2011).*

Delfino y otros (2011) ya mencionan en esa fecha a la Biblioteca trabajando activamente con las editoriales para desarrollar estándares y flujos de trabajo para el depósito de contenidos de publicaciones seriadas en NLM DTD y otros formatos. Mencionan además que para el proceso de depósito obligatorio de LC existen formatos preferidos por LC para las publicaciones seriadas electrónicas a ser depositadas (así como para otros materiales, tanto analógicos como digitales). Los formatos, en orden de preferencia son JATS y otros formatos XML estándar en primer lugar, seguidos de PDF, HTML, y luego todos los demás formatos. Destacan que la elección de formatos de preferencia refleja la evaluación que ha realizado la Biblioteca acerca de los formatos que mejor satisfacen sus objetivos de usabilidad para proporcionar el acceso a los investigadores actuales, y la capacidad de preservación de los mismos para garantizar el acceso de los investigadores en el futuro. Forma parte de la política general de la Biblioteca en cuanto a *e-serials* dar prioridad a la preservación del contenido sobre el aspecto de presentación del original.

Teniendo en cuenta lo mencionado en el párrafo anterior, los criterios para la adquisición de contenidos de publicaciones seriadas que aparecen en la estrategia de adquisiciones de LC del año 2011 son, en orden de prioridad: 1) adquirir ficheros XML (preferentemente JATS) directamente de las editoriales; 2) adquirir ficheros en un número limitado de formatos estándar de terceros, para reducir la variabilidad en los procedimientos de ingreso.

Hyde (2015) en su reporte sobre la necesidad de determinar una estrategia para llevar a cabo los depósitos y gestionar las colecciones digitales indica que el progreso del programa *eDeposit* ha sido poco claro y la auditoría llevada a cabo por la oficina del director general (OIG) no pudo determinar si el progreso del programa *eDeposit* ha cumplido con las expectativas de la dirección de la biblioteca en cuanto a costos, tiempo de ejecución y rendimiento. El Proyecto *eSerials* fue el único proyecto iniciado en el marco del Programa *eDeposit* desde que el programa fue creado en 2010, con un presupuesto estimado de \$ 11.830.000 para completar el mencionado proyecto *eSerials*. El presupuesto incluye una estimación para los contratistas que proveen el desarrollo de la infraestructura de tecnológica necesaria y la configuración de hardware y software, desde el año fiscal 2010 al 2015.

*By one measure, as of December 31, 2014, the Copyright Office was ingesting 835 electronic serial titles on an on-going basis as part of the eSerials Project. This represents approximately*

*34% of all the active eSerial scholarly titles published in the United States and available online (Hyde, 2015).*

En lo que esta auditoria sí identificó progresos fue en los acuerdos establecidos para *eSerials*, los cuales fueron de dos tipos: con proveedores de servicios editoriales y con editoriales. La Biblioteca y la Oficina de Derechos de Autor han establecido un acuerdo de cooperación con *Data Conversion Laboratory* (DCL). El acuerdo habilita a DCL a depositar publicaciones seriadas en respuesta a las demandas de depósitos obligatorios en nombre de sus editores, sin costo alguno para la Biblioteca. También han firmado acuerdos con tres editoriales: Walter de Gruyter, Emerald Publishing y Hindawi.

## Scholars Portal

*Scholars Portal* (SP) es un servicio de *Ontario Council of University Libraries* a disposición de profesores y estudiantes en 21 universidades de Ontario en Canadá. Es un ejemplo de implementación exitosa de JATS que comienza para las revistas, brindando servicios de alojamiento a un pequeño número de consorcios regionales, permitiéndoles también organizar y proporcionar sus propios servicios para el acceso y la preservación de revistas electrónicas.

*The primary purpose of the portal is access, but the consortium has made an explicit commitment to the long-term preservation of the e-journal content it loads locally. The initiative began with grant funding but as of 2006 became self-funded through tiered membership fees (Kenney y otros, 2006).*

SP implanta con éxito desde el año 2006 *NLM Journal Archiving and Interchange Tag Set* para su colección *SP Ejournal* basada en XML. *SP EJournal* maneja datos de 28 proveedores, de los cuales 20 de ellos implementan XML con *NLM Tag Suite* o *JATS XML*.

Tras el éxito de la adopción JATS en el repositorio SP de revistas, el equipo de SP decidió adoptar BITS como formato estándar para el repositorio SP de libros (Zhao y otros, 2015). Actualmente utilizan la extensión de JATS denominada BITS para los libros electrónicos de su colección *SP Ebook*. La plataforma *Scholars Portal* para libros electrónicos está diseñada para proporcionar una única interfaz para acceder a textos digitales de múltiples proveedores de contenido. El portal funciona como un agregador (*aggregator*) de contenidos de libros de varios proveedores y para ello convierten todo a un formato único utilizando BITS.

Zhao y otros (2015), explican las razones por las cuales adoptaron BITS: la primera es que BITS es un modelo XML para libros de las áreas STM, por lo tanto, el alcance de BITS es adecuado para la colección *SP Ebook*. En segundo lugar, BITS proporciona flexibilidad en la gestión de libros y partes de libros, permitiendo manejar éstas últimas como archivos separados, lo cual permite montar un documento final según se necesite. De esta forma, los libros y las partes de libros se pueden gestionar como archivos independientes y ser incluidos según sea necesario en el documento final. Por último, y destacada como la más importante de las razones, al ser BITS un modelo para libros basado en el modelo para artículos JATS, y debido a la experiencia previa adquirida con JATS, adoptar BITS minimiza el tiempo de aprendizaje del personal y permite reutilizar el sistema ya existente, desarrollado en base a JATS.

En este caso, no se utiliza BITS para producción editorial, sino para presentar un producto, los libros o partes de libros que adquieren a diversos proveedores. Hacen uso de la capacidad de intercambio de BITS para unificar el XML originario y de esa forma pueden “agregar” XML de diversas fuentes (editoriales) y crear una base unificada con todos ellos. De allí parte la versatilidad de los productos y servicios que *SP Ebook* puede brindar.

#### 6.4.2. Servicios editoriales

El desarrollo continuo de las Tecnologías de Información y Comunicación (TIC) en las actividades de la biblioteca y en las tareas de investigación estimula las iniciativas y las asociaciones entre la academia y la biblioteca. Dentro de éstas iniciativas están las vinculadas al área de la publicación, entre otros productos, de revistas digitales dirigidas por académicos. Actualmente, el papel de las bibliotecas académicas como editoriales es una realidad en muchas instituciones. Se dan algunos casos en que las bibliotecas ofrecen servicios de publicación digital de revistas en colaboración con las editoriales universitarias e incluso con editoriales particulares. Un indicio de la madurez de esta área profesional en las bibliotecas es la creación de *Library Publishing Coalition* (LPC). LPC es una asociación independiente cuyo propósito es apoyar las prácticas editoriales de las bibliotecas y promover los intereses de las bibliotecas que participan en actividades de publicación en sus campus.

*Library Publishing Coalition* define los servicios editoriales de la biblioteca, *Library Publishing Services* (LPS) como el conjunto de actividades dirigidas por bibliotecas de facultades y universidades para apoyar la creación, difusión y preservación de obras académicas, creativas y educativas. En general, los servicios editoriales de la biblioteca requieren un proceso de producción, se publican trabajos originales que no han estado disponibles anteriormente, y se aplica un nivel de certificación de los contenidos publicados, sea a través de la revisión por pares o de la extensión de la marca institucional. Sobre la base de los valores fundamentales de la biblioteca y de las habilidades tradicionales de los bibliotecarios, se distingue de otros campos de publicación en la preferencia por la difusión en acceso abierto de los contenidos publicados y la voluntad de acoger además otros modos informales y experimentales de la comunicación académica, y desafiar el *status quo* (Lippincott, 2014).

Los servicios editoriales de las bibliotecas se diferencian del trabajo de otras editoriales (incluidas las editoriales comerciales, las de las sociedades científicas y las académicas) en gran parte por su modelo de negocio, que muy frecuentemente se basa en ser subvencionado por el presupuesto de la biblioteca, en lugar de operar principalmente como actividad con fines de lucro o para recuperar los costos. Las bibliotecas son relativamente recién llegadas en este ámbito editorial, habiendo incursionando en este trabajo en el entorno digital en los últimos 20 años (Skinner y otros, 2014).

*Publishing services are rapidly becoming a norm for research libraries, particularly journal publishing services [...] the question is no longer whether libraries should offer publishing services, but what kinds of services libraries will offer* (Hahn, 2008).

Hahn (2008), en un reporte para la *Association of Research Libraries* (ARL) de Estados Unidos, describe a los servicios editoriales en las bibliotecas como formando parte de una gama de nuevos tipos servicios que las bibliotecas han desarrollado o están desarrollando, y que son co-administrados y a menudo integrados con esos otros nuevos servicios, entre los que incluye las iniciativas de digitalización, las iniciativas en el área de las humanidades digitales, los repositorios digitales, así como también el desarrollo de objetos de aprendizaje y las actividades de preservación digital.

Roseblum (2008) manifiesta que, aunque la actividad editorial de las bibliotecas es sólo una pequeña fracción comparada con el amplio mundo de las editoriales académicas, estas actividades llenan una necesidad importante, proporcionando apoyo y visibilidad a nuevas revistas electrónicas, o a revistas producidas por los departamentos que carezcan de personal y recursos para publicar en línea. Además, desempeñan un papel muy útil en el asesoramiento a los editores de revistas y editoriales universitarias que están tratando de entender los cambios en la publicación académica y la gama de opciones de publicación disponibles para ellos.

Una de las principales razones de que se hayan iniciado servicios editoriales en las bibliotecas es la de dar respuesta a la insatisfacción con las editoriales tradicionales que con frecuencia pretenden la obtención de derechos casi exclusivos de los trabajos que publican, además de crear productos caros, con lo cual dificultan a los autores el uso de sus propios trabajos y a su vez impiden el acceso amplio y asequible a los lectores. En respuesta a esto, los servicios editoriales de las bibliotecas se han puesto el objetivo de publicar de forma más rentable, proporcionando a su vez términos más justos a los autores que las editoriales tradicionales, especialmente al permitir a los autores conservar los derechos de autor, garantizando a la editorial sólo aquellos derechos necesarios para la publicación (Hawkins, 2015).

Lorimer (2013), refiriéndose a la forma en que varias organizaciones de Canadá están proporcionando servicios de publicación profesionales orientados a revistas, menciona que la mayoría de los servicios editoriales prestados por las bibliotecas ofrecen un sistema en línea y el correspondiente soporte técnico, mientras que otras actividades relacionadas con la edición, tales como los procesos de revisión de artículos, correcciones de estilo, y la gestión de suscripción o membresía siguen siendo responsabilidad de cada revista. Mattson y Friend (2014) en cambio utilizan el concepto de "niveles de servicio", relacionándolo con el grado en que la biblioteca y los editores están involucrados en la prestación de servicios de publicación. En los procesos de negociación e implementación que se llevan a cabo entre las bibliotecas y los editores de las revistas, ambas partes negocian los niveles en base a los roles y expectativas de cada uno.

La Coalición para la Información en Red (*Network Information Coalition*) menciona que en 2012 el modelo editorial de las bibliotecas incluía la publicación en digital a través de plataformas de alojamiento, más comúnmente *Open Journal Systems* (Missingham y Kanellopoulos, 2014).

En la edición digital, la codificación de los documentos en XML puede significar una serie de ventajas para las bibliotecas que han invertido en *hosting* y servicios editoriales o planean hacerlo en el futuro. Los flujos de trabajo XML permiten, además de obtener varios formatos de despliegue de forma rápida y sencilla (ePub, HTML, PDF), reutilizar el contenido en catálogos, en sitios web, en bases de datos, generar productos impresos, automatizar los procesos, escalar servicios, y preservar el contenido digital para el futuro (Mcglone, 2013).

Como ya se introdujo en la sección 6.1, existen diferentes formas de trabajo editorial con XML. Muchos servicios editoriales de bibliotecas utilizan actualmente la modalidad *XML-out*, y lo fundamentan en el tipo de servicios que brindan. Mcglone (2013) refiriéndose al caso de *Michigan Publishing* (producto de la fusión de los servicios editoriales de la biblioteca con la editorial de la universidad) menciona que ellos utilizan JATS XML, Scribe y TEI, además de un flujo de trabajo generalizado para la creación y publicación de documentos codificados en XML para monografías y artículos de revista que comienza con un documento de Word, y termina como un documento marcado con XML, para presentar luego el artículo o monografía a los lectores en HTML y PDF, estando también disponible para reutilización con otros propósitos.

*The digital publishing activities of Michigan Publishing have relied primarily on XMLOut because of our tendency to provide only conversion and hosting services for journals and digital monographs, similar to many libraries providing publishing services* (Mcglone, 2013).

Hasta hace poco, los servicios editoriales en bibliotecas principalmente se han vinculado con las bibliotecas de investigación, en particular con aquellas asociadas con las editoriales universitarias. Sin embargo, los programas editoriales en pequeñas bibliotecas universitarias ofrecen un modelo alternativo, pues estos hacen la diferencia al enfatizar la enseñanza, el aprendizaje y la investigación a



nivel de grado, no sólo con la publicación de revistas editadas por los estudiantes, sino también por la formación de los estudiantes para realizar esa labor. En lugar de optar por una instalación propia de software de código abierto, las bibliotecas académicas pequeñas suelen utilizar soluciones de hosting, pues esto les permite centrarse en los servicios. Están limitadas por presupuestos más ajustados, por lo cual generalmente carecen de un puesto a tiempo completo para el trabajo editorial. La opción para ellas es asignar un porcentaje de tiempo del personal existente, aprovechando la experiencia su personal para proporcionar servicios editoriales sin necesidad de contratar personal adicional, pero además desarrollando políticas que les permiten establecer límites y expectativas, y trabajando en estrecha colaboración con los profesores, estudiantes y unidades del campus para avanzar en objetivos comunes (Spiro, 2015).

Aunque labor editorial es compatible con las habilidades tradicionales del bibliotecario, hay habilidades adicionales que éstos deben dominar si quieren proporcionar servicios editoriales robustos a la comunidad académica. Según Skinner y otros (2014), y refiriéndose a la realidad norteamericana, la mayoría de los programas educativos existentes relacionados con edición y publicación se centran en la producción de publicaciones a nivel comercial, dejando un vacío en la formación para publicar en áreas académicas y científicas. Y como existen varios sub-campos dentro de la edición (comercial, académica, STM, es decir científica, técnica y médica), donde cada uno de ellos requiere un conjunto diferente de cualificaciones, esto presenta un desafío para la existencia y el crecimiento de programas de formación en esta área.

La gama de habilidades tecnológicas que son útiles para la tarea editorial incluidas en la formación bibliotecaria, se aprenden con la posibilidad de tomar cursos en lenguajes de marcado como XML, lenguajes de script, PHP y SQL para aplicaciones de bases de datos, etc. Esas habilidades son cada vez más importantes para publicar en digital y hacer disponibles las publicaciones académicas en línea. Otra ventaja de la formación bibliotecaria al momento de aportar habilidades y competencias profesionales al ámbito editorial son los conocimientos sobre preservación digital, otro campo cada vez más vigente, pues las obras originales están naciendo digitalmente y los requisitos necesarios para preservarlas son más complejos. Mientras que la preservación puede no representar una parte típica del proceso de publicación digital, los metadatos sí, y éstos son parte integral de la preservación digital. La creación y gestión de metadatos es también una habilidad tradicional de los bibliotecarios (Keener, 2014).

En el siguiente capítulo se explora si las bibliotecas académicas están utilizando XML en sus procesos editoriales. Es de particular interés recoger información sobre si las bibliotecas que brindan servicios editoriales implantan actualmente JATS o no, si lo utilizan como lo implementan y si no lo hacen cuales son los motivos. Para ello se recurre a una encuesta en línea, destinada a bibliotecas que efectivamente prestan estos servicios.



## CAPITULO 7. ENCUESTA A BIBLIOTECAS QUE OFRECEN SERVICIOS EDITORIALES

La edición 2016 de *Library Publishing Directory* se convierte en la fuente principal de contactos a la hora de distribuir la encuesta para conocer la situación actual de implementación de JATS en los servicios editoriales ofrecidos por bibliotecas en la actualidad. LPD 2016 (publicado en octubre de 2015), cuenta con contactos correspondientes a 115 bibliotecas, 105 de América del Norte (Estados Unidos y Canadá) y las otras 10 de Australia, Reino Unido, Suecia y Rumania.

*Library Publishing Directory* se publica anualmente desde 2013. El directorio 2016 ha sido elaborado consultando las bibliotecas que participaron en la edición anterior y además se adicionaron nuevas entradas por medio de la difusión de un cuestionario en listas de correo y redes sociales. Se ha contactado a las bibliotecas participantes de este directorio pues es una fuente concreta que cuenta con contactos recientemente recopilados y actualizados, pertenecientes a bibliotecas académicas de prestigiosas instituciones que efectivamente están brindando servicios de publicación.

La encuesta ha sido distribuida a través del correo electrónico a cada uno de los contactos que aparecen en el directorio para estas 115 bibliotecas, de manera personalizada, explicando los motivos de la encuesta, cómo serán tratados los datos y con qué fines serán utilizados los resultados. El formulario en línea de la encuesta permaneció abierto a la recepción de respuestas durante un período de 30 días desde mediados de noviembre a mediados de diciembre de 2015. El formulario empleado se incluye en el Anexo 1.

El diseño del formulario y el texto de las preguntas están centrados en el interés por saber si esas bibliotecas implementan JATS, y cómo lo hacen. Pero aún en el caso que no lo utilicen, será útil saber cuál es la opción que han escogido. Es por eso que no se pregunta específicamente por JATS (aunque sí se lo mencione como ejemplo junto a otros formatos en el texto de la pregunta 4).

La recepción de la encuesta ha sido buena teniendo en cuenta que de las 115 bibliotecas objeto de esta investigación, se obtuvieron un total de 53 respuestas, 45 de ellas mediante el formulario en línea, y 8 respuestas recibidas vía correo electrónico.

Previo a la presentación y el análisis de los resultados, presentamos un resumen del contexto de las bibliotecas objeto de esta encuesta.

### 7.1. Contexto de las bibliotecas objeto de la encuesta

A los efectos de complementar la información que estamos recogiendo en la encuesta, y para contar con información complementaria que aporte al análisis y la discusión de los resultados de la misma, se han tomado algunos datos estadísticos aportados por el directorio 2016, especialmente aquellos datos que dan a conocer la situación de las bibliotecas del directorio respecto a la variedad y el volumen de sus publicaciones, al personal dedicado a la tarea y a las tecnologías relacionadas con los servicios editoriales ofrecidos.

Las bibliotecas listadas en el directorio 2016 han publicado un total de 404 títulos de revistas académicas (*faculty-driven, campus-based journal titles*) durante el período junio 2014 - julio 2015. Y a esto se le suman los títulos de revistas gestionadas por los estudiantes, pues también se publicaron 224 títulos de revistas editadas por estudiantes en ese mismo período. En promedio, actualmente cada

biblioteca de este directorio publica aproximadamente unas cinco revistas, teniendo en cuenta las gestionadas por los académicos del campus, más las gestionadas por los estudiantes.

Según LPD edición 2016, las bibliotecas reportaron un total de 773 monografías publicadas en 2014 -2015. Tres bibliotecas- *California Digital Library*, *Cornell University Library*, e *Indiana University Library* publicaron 478 de esos títulos.

También de los datos aportados por LPD 2016 se conoce que las bibliotecas mayoritariamente financian sus servicios editoriales parcial o totalmente por el presupuesto de funcionamiento de la biblioteca, una estadística que en opinión de LPC afirma el posicionamiento de los servicios editoriales como un servicio de la biblioteca.

Las bibliotecas encuestadas utilizan diversas tecnologías para prestar los servicios editoriales, desde plataformas de código abierto a plataformas comerciales y también desarrollos locales de software propio. El 37% de las bibliotecas del directorio utiliza sólo una plataforma o software de publicación, mientras que la gran mayoría utiliza múltiples plataformas y otras herramientas.

*Popular publishing platforms include OJS (43%) and bepress (42%), followed by DSpace (31%), WordPress (25%), locally developed software (23%), CONTENTdm (19%), Fedora (16%), and OCS, OMP, EPrints, Islandora, Hydra, Scalar, Drupal, Tizra, Omeka, Biblioboard, and DPubS (each with less than 10% usage)* (Lippincott, 2015b).

Estos porcentajes resultan muy significativos para el análisis de la realidad actual de implantación de JATS XML, u otros lenguajes basados en XML, pues las plataformas de publicación (software utilizado) condicionan los flujos de trabajo. OJS no incluye herramientas para integrar XML en su sistema editorial y es utilizado por el 43 % de las bibliotecas que integran el directorio.

El promedio de personal profesional de todas las bibliotecas incluidas en la edición 2016 de LPD es de 2,4, registrándose un incremento de 1,8 respecto a la edición 2015 del directorio. Las bibliotecas también reportan un promedio de una persona a tiempo completo en la categoría de personal auxiliar, 0,8 en apoyo de estudiantes de posgrado, y una en apoyo de estudiantes de pregrado a sus servicios editoriales. Aproximadamente la mitad de las bibliotecas emplean personal de apoyo para sus programas editoriales. El 19% de las bibliotecas del directorio tienen empleados que son estudiantes de posgrado y un 30% emplean estudiantes de pregrado.

Las actividades editoriales de la biblioteca se encuentran insertas dentro de diversas secciones de la biblioteca, y son asumidas por profesionales en diferentes niveles de carrera, con títulos como bibliotecario de comunicación académica (*Scholarly Communication Librarian*), bibliotecario académico digital (*Digital Scholarship Librarian*, *Head of Digital Scholarship*), director de publicación digital (*Director of Digital Publishing*) o bibliotecario de iniciativas digitales (*Digital Initiatives Librarian*).

Los integrantes de LPC también han señalado en varias oportunidades la relativa falta de educación continua dirigida a aquellos que están participando en las tareas publicación, ya sea en las bibliotecas, las editoriales universitarias, o en el entorno de la publicación comercial.

## 7.2. Presentación y análisis de las respuestas recibidas por correo electrónico

Las respuestas recibidas vía correo electrónico son breves, y aportan menos información de la que se pretendía con el formulario, pero aun así constituyen un insumo valioso para el análisis. Son ocho las bibliotecas que si bien no envían respuestas a través del formulario como se solicitaba, hacen llegar la siguiente información respondiendo directamente al correo-e:

*"We don't offer publishing services for journals here at BYU. We host journals, but we don't run publishing services. Nor do we use XML. So, I wouldn't be a good candidate for this survey."*

R01

*"We are currently reorganizing our publishing services, and may implement xml in the future, but I honestly don't know."*

R02

*"Note, our library does not utilize a journal publishing system. We are a repository."*

R03

*"Our repository doesn't currently support XML, so we have never had a need to learn or use it as far as our publishing format goes. That may change and we'll cross that bridge when it comes :-)"*

R04

*"Our efforts so far are ad hoc and so we don't actually have a workflow - we've published one monograph and digitized some local materials but do not publish journals here."*

R05

*"I'm not familiar with using XML for publishing services."*

R06

*"XML? It hasn't come up - we're barely getting started."*

R07

*"I am sorry for the delay, but I did want to let you know that we are not using XML as you ask below for our publishing services."*

R08

Además de esos 8, recibimos también dos correos electrónicos cuyos remitentes manifiestan que han participado de la encuesta con envío de información a través del formulario, pero a su vez agregan más información:

*"Great topic! I'd love to stay in touch with you as you work on this project. I'm a big proponent of XML-centric workflows, though the tooling is lacking for their implementation at present. I hope we can change that in the future."*

R09

*"I had a hard time completing this survey accurately since the answers are different depending on the context. My colleagues and I define digital publishing broadly to include "traditional" digital collections, content in IRs, e-books and e-journals (hosted and local open source, OA and hybrid), as well as finding aids. The survey did not provide a way to represent more than one "stream" so I answered the question with a more narrow hosted e-journal focus. If I broadened the context for #4 to include our open source e-journal platform, the answer would be JATS in the near future; if extended even further to include special collections finding aids, the answer would include EAD. For #5, the respective answers would be "at the end..." for the e-journal context and "all the way through..." for the finding aids. I hope this makes some sense to you. Best of luck with your research."*

R10

Las respuestas R01 a R08 provienen de bibliotecas que consideran no estar ofreciendo servicios editoriales o bien que no están utilizando XML para los mismos. Algunas incluso expresan no saber si lo van a utilizar en un futuro. En ambos casos son bibliotecas que no están utilizando actualmente XML en sus procesos editoriales. Los motivos que se explicitan en las respuestas se pueden agrupar en tres variantes: 1) los que no están seguros si implantarán XML o no en el futuro (R02); 2) los que se limitan a contestar que actualmente no lo utilizan (por varias razones: porque están recién incursionando en DPS (como R05 y R07), o porque no publican revistas sino que las alojan (conciben un *journal publishing system* diferente de un *repository* como R03 y R04); 3) los que manifiestan no estar familiarizados con el uso de XML en DPS (R06).

Para R01 y R03 existe una distinción clara entre alojar documentos y prestar un servicio editorial. R10 y sus colegas en cambio manejan una definición amplia cuando se refieren a *digital publishing*, y así lo explicitan mencionando que incluyen colecciones digitales “tradicionales” en dicho concepto. No obstante, la definición de LPS que maneja LPC es bastante amplia como para incluir dentro de DPS a los repositorios institucionales, con lo cual se justifica que el directorio LPD incluya bibliotecas que solo publican *Electronic Theses and Dissertations* (ETDs).

De la respuesta de R09 se pueden sintetizar tres cosas: una apuesta fuerte a nivel personal a favor del modelo de publicación centrado en flujos de trabajo con XML, las debilidades que en opinión del entrevistado tiene actualmente el modelo centrado en XML en cuanto a carencias en las herramientas para su implementación, y la esperanza de hacer algo al respecto para cambiar esa realidad a futuro.

En el servicio por el cual responde R10 se publican distintos tipos de documentos y tienen varios flujos de trabajo establecidos. En este caso y otros donde la producción editorial de la biblioteca es más diversa, existiendo un flujo de trabajo diferente para publicar cada tipo de documento, algunas preguntas del formulario resultan demasiado específicas para revistas. Esto se debe básicamente a la orientación temática de este Trabajo Fin de Máster con foco en las revistas científicas.

### 7.3. Análisis y discusión de los resultados de encuesta

El cuestionario de la encuesta cuenta con cinco preguntas diseñadas para recoger fundamentalmente si las bibliotecas utilizan XML o no en los procesos de publicación, qué lenguaje, aplicación o vocabulario utilizan, y cómo lo implementan. La primera pregunta está dirigida a saber si utilizan XML actualmente, si piensan utilizarlo a futuro o si no lo implementan. En caso de las que no implementan ni piensan implementar XML, se les pregunta cuáles son los motivos por los cuales no lo hacen. A todas las bibliotecas se les pregunta su opinión respecto a las ventajas de utilizar XML en los procesos editoriales. La quinta y última pregunta está dirigida a aquellas bibliotecas que han respondido afirmativamente respecto a la implantación de XML en sus procesos editoriales y también a las que han manifestado que piensan implementarlo en el futuro. Lo que se les pide en esta pregunta es que especifiquen en que momento del proceso incorporan XML.

A continuación, presentamos el análisis y la discusión de la información contenida en cada una de las respuestas recibidas mediante el formulario en línea. El análisis está encabezado por la pregunta correspondiente en su versión en español.

**Pregunta 1. ¿Están implementando XML actualmente en las actividades editoriales de su biblioteca?**

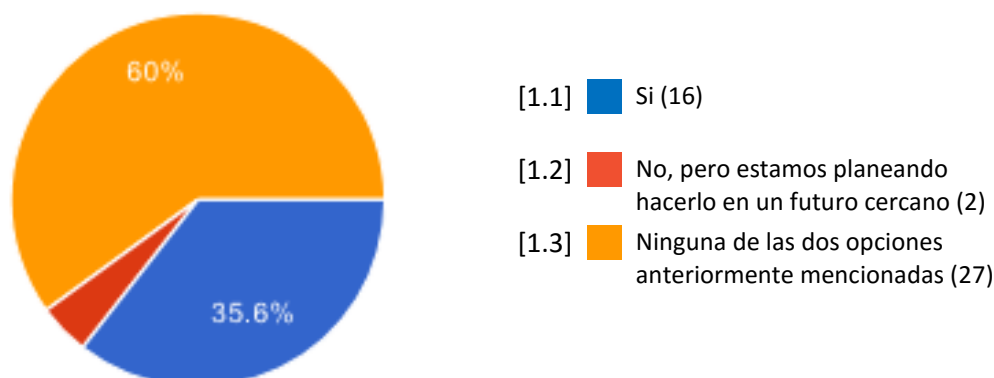


Gráfico 1. Implementación de XML en LPS  
Fuente: elaboración propia en base a los resultados de la encuesta.

La mayoría de las bibliotecas que han respondido mediante el formulario (60%), manifiestan no utilizar XML en sus procesos editoriales. Si considerásemos además como respuesta indirecta para la opción “Ninguna de las dos opciones anteriores” (codificada como 1.3) la de esas ocho bibliotecas que enviaron sus respuestas por correo electrónico, entonces correspondería a un 66% el porcentaje de las bibliotecas que no utilizan actualmente XML en sus procesos editoriales.

El elevado porcentaje de bibliotecas que no utilizan XML coincide con los datos de LPD respecto a las plataformas de publicación que utilizan. Que OJS no incluya herramientas para trabajar de forma interna con XML y sea utilizado por el 43% del total de las bibliotecas que figuran en LPD edición 2015 es muy significativo al momento de corroborar esta tendencia; la segunda plataforma de publicación de contenidos más utilizada es *bepress* y la tercera *DSpace*. Respecto a estas dos últimas, ambas utilizan XML solo para los metadatos de los documentos que alojan.

Como ya se ha mencionado, existen diferencias entre las bibliotecas respecto al “nivel de servicio” que negocian biblioteca y editores de cada revista. Los servicios que ofrecen pueden ser de alojamiento (fundamentalmente repositorios y plataformas de publicación de contenidos que no incluyen XML en sus procesos) o servicios editoriales que implican editar y componer contenidos en texto completo y pueden llegar a integrar XML en el proceso o en alguna etapa del proceso. Depende del rol que asuma cada biblioteca. Muchas de ellas no implantan XML pues no lo necesitan.

**Pregunta 2. Si su respuesta a la primera pregunta es "ninguna de las dos opciones anteriores", por favor especifique los motivos eligiendo una o más de las siguientes opciones para completar esta afirmación: No estamos interesados en la aplicación de XML en nuestras actividades editoriales porque...**

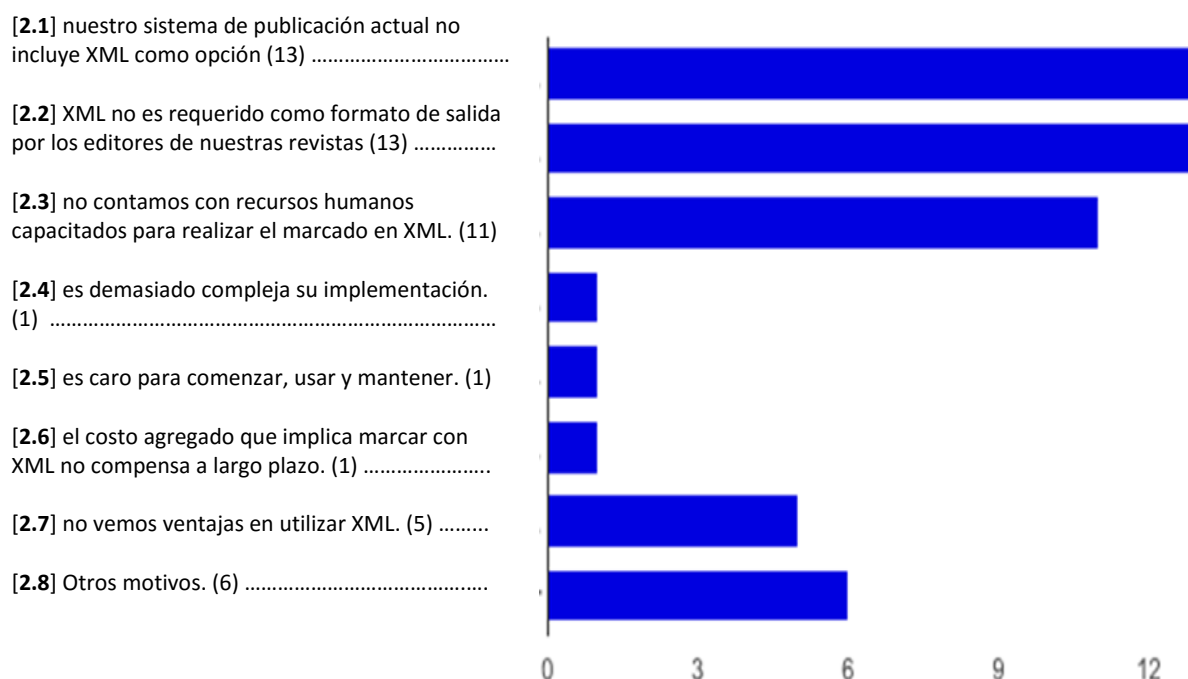


Gráfico 2. Motivos por los cuales no implementan XML en LPS  
Fuente: elaboración propia en base a los resultados de la encuesta.

Se obtuvieron un total de 27 respuestas a esta pregunta, para la cual se podían seleccionar una o varias de las opciones de respuesta que aparecen en el formulario (codificadas 2.1 al 2.7) y además escribir opciones propias de respuesta en "Otros motivos" (codificada como 2.8). Los textos de las respuestas a la opción 2.8 están disponibles en ANEXO 2, Textos correspondientes a las preguntas con opción a respuestas abiertas.

El número de respuestas a esta pregunta se corresponde con el número de encuestados que respondieron vía formulario marcando la opción "Ninguna de las dos opciones anteriormente mencionadas" para la **Pregunta 1**. Dicho de otro modo, todos los que respondieron que no utilizan actualmente XML para DPS y que no están pensando en utilizarlo en un futuro, han contestado a la **Pregunta 2** para expresar las razones por las cuales no lo están utilizando. Las principales razones que se recogen en las respuestas son, en el primer lugar con la mayoría de respuestas, dos motivos principales: 1) que el sistema de publicación utilizado actualmente por la biblioteca no incluye la opción XML (codificada como opción 2.1) y 2) que XML no es un formato requerido por los editores de las revistas a las que ellos ofrecen servicios editoriales (codificada como opción 2.2). El motivo que aparece en segundo lugar es que no cuentan con recursos humanos capacitados para llevar a cabo la tarea de marcado con XML (codificada como opción 2.3).

Ocho de las trece bibliotecas que indicaron los motivos 1) y 2), incluyeron ambos motivos conjuntamente entre sus opciones de respuesta (EN03, EN12, EN13, EN35, EN26, EN38, EN43, EN44).



Solo tres de ellas marcaron ambos motivos como las únicas razones por las cuales no implementan XML (EN12, EN25, EN43).

De las once bibliotecas que incluyeron como motivo no contar con recursos humanos capacitados para realizar el marcado en XML (codificada como opción 2.3), solo 3 marcaron este motivo como única razón (EN04, EN23, EN35).

Cinco de las bibliotecas que respondieron la encuesta no ven ventajas en utilizar XML. Dos de ellas señalan esa como la única razón para no implantar XML en DPS (EN09 y EN11), mientras que las tres restantes indican además otras dos razones: 2.1 y 2.2 (EN02), 2.2 y 2.8 (EN19), 2.2 y 2.3 (EN32).

Otros motivos que influyen en la decisión de no utilizar XML, pero en menos casos, refieren a que lo perciben como un sistema de publicación complejo de implementar (EN29), caro de implantar y mantener (EN26), o porque consideran que los costos asociados a marcar con XML no compensan a largo plazo (EN29).

Seis bibliotecas que contestaron tener “Otros motivos” diferentes de las opciones que aparecen en el formulario, no utilizan XML porque están conformes con su sistema de publicación actual (EN02), o porque tienen otras prioridades (EN05), o porque no resulta útil para publicar multimedia (EN19). Una de las bibliotecas dice que no lo está usando porque no ha llegado el momento (EN38). Dos de ellas no mencionan en realidad otros motivos en esta opción, sino que expresan otras cosas: una aporta información sobre sus experiencias ayudado usuarios que requerían esos servicios y les ha brindado información de contacto de empresas que realizan el marcado (EN39) y otra aclara que podrían llegar a implementarlo en el futuro (EN45). Esta última biblioteca solo ha contestado a la **Pregunta 2** indicando tener ese motivo, las otras bibliotecas, además de expresar “Otros motivos”, incluyen también opciones predeterminadas del formulario en su respuesta.

**Pregunta 3. En su opinión, ¿cuáles son las ventajas de la adopción de XML (aplicación, vocabulario o lenguaje) en los servicios editoriales brindados por las bibliotecas?**

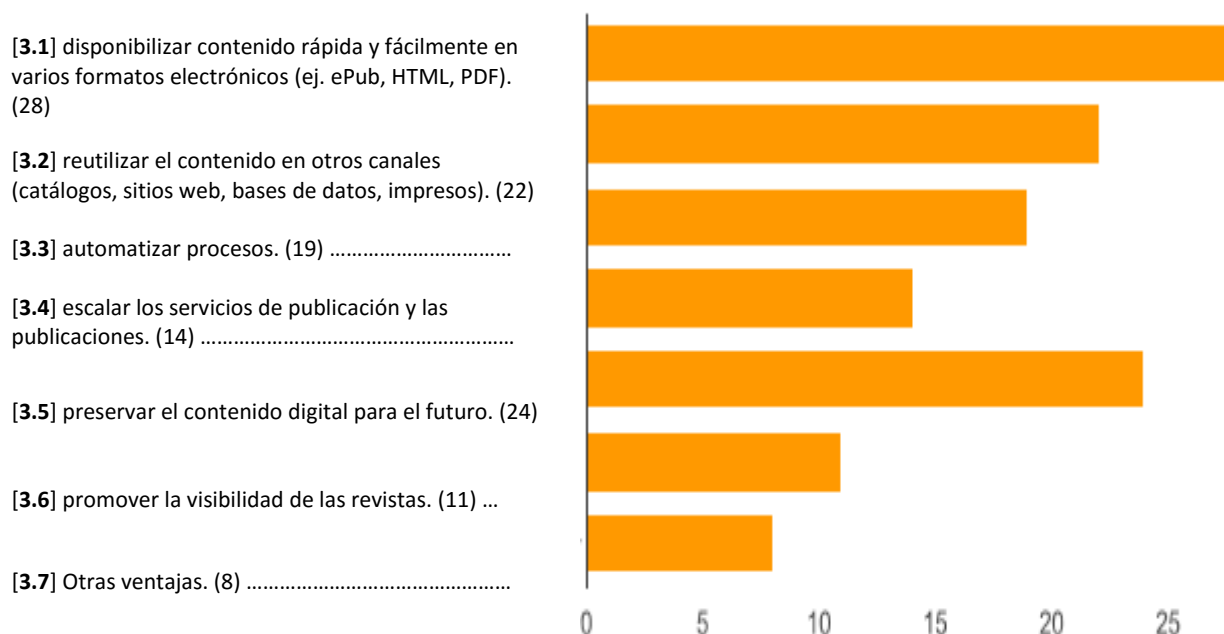


Gráfico 3. Ventajas de implementar XML en LPS  
Fuente: elaboración propia en base a los resultados de la encuesta.

De las 45 bibliotecas que respondieron mediante el formulario, 40 respondieron a esta pregunta, para la cual se podían seleccionar una o varias de las opciones de respuesta ofrecidas en el formulario (codificadas del 3.1 al 3.6) y además escribir opciones propias de respuesta (codificada como 3.7). Excepto por tres casos en los que respondieron no estar seguros o no conocer lo suficiente del tema para tener una opinión, todos los demás encuestados encuentran ventajas al momento de utilizar XML. Eso resulta muy significativo, más teniendo en cuenta que el 66% no utiliza XML en las actividades de su biblioteca (resultado de la **Pregunta 1**). En otras palabras, la gran mayoría de los que respondieron no estar utilizándolo actualmente, reconoce ventajas en XML.

Las ventajas que más mencionan los encuestados a la hora de implementar XML en DPS, en orden creciente según la cantidad de respuestas obtenidas, son: disponibilizar contenido rápida y fácilmente en varios formatos electrónicos (ej. ePub, HTML, PDF) (28), preservar el contenido digital para el futuro (24), reutilizar el contenido en otros canales (catálogos, sitios web, bases de datos, impresos) (22) y automatizar procesos (19). En menor medida mencionan otras ventajas como escalar los servicios de publicación y las publicaciones (14) y promover la visibilidad de las revistas (11).

Seis encuestados escribieron sus propias respuestas, diferentes de las opciones que brinda el formulario. De las respuestas abiertas que consignaron en "Otras ventajas" (disponibles en ANEXO 2, Textos correspondientes a las preguntas con opción a respuestas abiertas), solo cuatro de ellas aportan información relevante: EN33 menciona como ventaja la cosecha de contenidos, EN34 la carga retrospectiva de registros por lotes, EN38 valora la ventaja de XML para el envío de contenidos a organizaciones vinculadas con el acceso abierto como PubMed y EN44 considera que es ventajoso

promover estándares para compartir datos. Las otras dos respuestas (EN32 y EN39), solo expresan desconocimiento del tema o inseguridad respecto a las ventajas.

**Pregunta 4. ¿Qué formato XML (aplicación, vocabulario o lenguaje) están utilizando / planificando utilizar y por qué? (Por ejemplo, JATS, TEI, DB).**

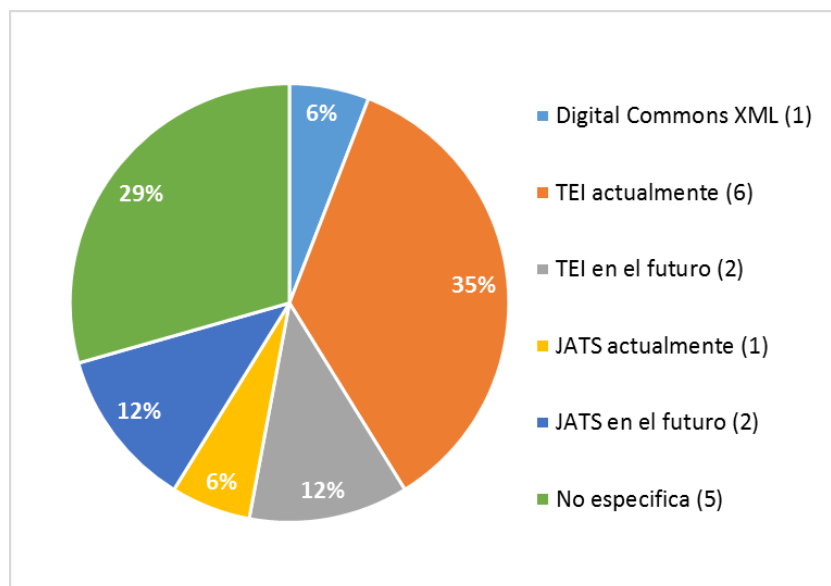


Gráfico 4. Formatos XML utilizados por las bibliotecas

Fuente: elaboración propia

Se obtuvieron diecisiete respuestas a la **Pregunta 4**, disponibles en el ANEXO 2, Textos correspondientes a las preguntas con opción a respuestas abiertas.

De las diecisiete bibliotecas que han respondido, nueve se inclinan por TEI: seis usan TEI actualmente (EN01, EN08, EN14, EN22, EN36, EN40), una lo utilizó en el pasado (EN34) y dos piensan hacerlo en el futuro (EN21, EN31). Lo utilizan tanto para libros como para otros proyectos de digitalización. Solamente una de las bibliotecas encuestadas está utilizando JATS actualmente (EN42), y otras dos planean comenzar a utilizar JATS en el futuro para nuevas revistas (EN08 y EN31). EN08 menciona que utilizan TEI actualmente, pero tienen previsto utilizar JATS para las nuevas revistas. Una biblioteca utiliza *Digital Commons XML* (EN28).

Cinco de las bibliotecas que han respondido a la **Pregunta 4** expresan que utilizan XML, pero no especifican qué lenguaje, vocabulario o aplicación emplean. EN41 menciona que generan XML automáticamente con *InDesign*<sup>19</sup>, por lo cual posiblemente generen XML a partir de archivos PDF. Una de las bibliotecas dice no conocer el lenguaje utiliza su proveedor de servicios (EN10) y otra no sabe aún que lenguaje implantarán (EN20). EN21 y EN37 tampoco mencionan qué lenguaje XML utilizan.

De las respuestas a la **Pregunta 4** también resulta que **seis de las bibliotecas no utilizan actualmente XML integrado a sus procesos editoriales**, sino para otros propósitos: dos de las bibliotecas aclaran que solo usan XML para importar documentos por lotes (EN15, EN34). EN34 también lo utiliza para indizar contenido en bases de datos. Otras dos de las bibliotecas manifiestan utilizar lenguajes XML solo para los metadatos, no para marcar el texto completo (EN21 y EN37), otra

<sup>19</sup> InDesign es una aplicación para la maquetación y el diseño desarrollada por Adobe Systems.

para conversión entre lenguajes (EN28), y la restante menciona específicamente que utiliza XML para exportar registros desde *Dublin Core* (DC) a otros formatos (EN33). Como las bibliotecas que han respondido a la **Pregunta 4** se corresponden con las que han respondido afirmativamente a utilizar o estar pensando utilizar XML para DPS en la **Pregunta 1**, excepto por una biblioteca que no ha contestado la **Pregunta 4** (EN17), podemos deducir que no todas las bibliotecas que han afirmado implementar XML para brindar servicios editoriales en la **Pregunta 1** lo están utilizando para esos fines.

En resumen, de las bibliotecas que utilizan XML para sus actividades (editoriales u otras), o piensan utilizarlo en el futuro, hay cinco bibliotecas que a pesar de manifestar que utilizan o utilizarán XML, desconocen o no saben aún qué lenguaje, aplicación o vocabulario utilizarán. De las doce que especificaron el lenguaje utilizado, solo una implanta JATS actualmente y dos piensan utilizarlo en el futuro, frente a seis que utilizan TEI actualmente y dos que piensan implantarlo en el futuro. Una sola biblioteca expresa utilizar *Digital Commons XML*.

#### Pregunta 5. ¿En qué punto integran/están planeando integrar XML en el flujo de trabajo?

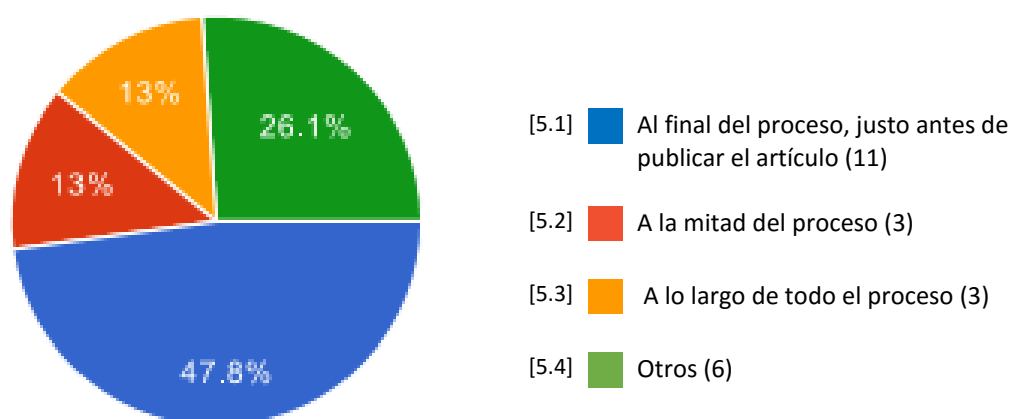


Gráfico 5. Forma de trabajo con XML en el proceso editorial - LPS  
Fuente: elaboración propia en base a los resultados de la encuesta

Se han obtenido 23 respuestas a esta pregunta, para la cual se podía seleccionar una de las opciones de respuesta ofrecidas en el formulario (codificadas 5.1 al 5.3), además de la posibilidad de escribir opciones propias de respuesta (codificada como 5.4). Los textos de las respuestas para la opción 5.4 están disponibles en ANEXO 2, Textos correspondientes a las preguntas con opción a respuestas abiertas.

La mayoría de las bibliotecas que respondieron esta pregunta opta por integrar XML al final del proceso editorial (47.8%), y solo un 13% utiliza XML durante todo el proceso. Otro 13% manifiesta incorporar XML a la mitad del proceso editorial. Incorporar XML al final es más sencillo y menos costoso de implantar. Como ya se ha mencionado en el capítulo 6, el modelo de negocio de las bibliotecas que brindan servicios editoriales es un modelo sin fines de lucro, que depende enteramente en la mayoría de los casos del presupuesto de la biblioteca y del personal de la misma. Además, teniendo en cuenta los datos de LPD sobre el número de títulos de revista que publican las bibliotecas objeto de esta encuesta (5 en promedio), es razonable que la opción más económica y más fácil de implantar resulte ser la más sea utilizada.

Para la categoría "Otros" se han recibido ocho respuestas con especificaciones y aclaraciones. Seis de los encuestados utilizaron solo esta opción para responder a la **Pregunta 5**, mientras que dos

de ellos (EN01 y EN22) además escogieron una de las opciones predefinidas de respuesta entre las tres de este tipo que se ofrecían en el formulario: “Durante todo el proceso editorial, los artículos son cargados al Sistema en XML”. EN01 aclara que ellos tienen flujos de trabajo diferentes para diferentes productos y EN22 explica que solo utilizan flujos de trabajo XML con TEI para los libros que publican, no así para otros productos como ETDs y revistas académicas en el repositorio (todas publicadas en PDF). En el caso de EN34, utilizan XML luego de publicar (parecería que se están refiriendo a la opción *XML-out*). Dos bibliotecas dicen que utilizan XML para describir contenidos en el repositorio (EN32 y EN40).

Tres de las respuestas consignadas en “Otros” son explicaciones de por qué no implementan flujos de trabajo XML: no tienen personal ni conocimientos para implementarlo (EN38), solo lo implantarían si se vuelve necesario / valioso (EN39), o planean integrarlo (EN29).

La pregunta 5 no ha resultado acertada para recoger información sobre formas de trabajo con XML en bibliotecas que ofrecen servicios editoriales para diferentes productos (revistas, libros y otros), pues no da opción de respuesta para quienes trabajan con varios flujos en sus procesos editoriales. El motivo es que la pregunta fue pensada teniendo en mente la producción de contenidos de revista. Como consecuencia, la mayoría de los que estaban en esta situación optaron por aclarar en “Otros” en el formulario o escribir un correo-e explicando la diversidad de flujos de trabajo que implementan.

### Comentarios en el formulario

Al final del formulario de la encuesta se agregó un cuadro de texto con la finalidad de recabar otra información extra que los encuestados pudieran aportar sobre aspectos de los procesos y flujos de trabajo editorial, mediante la inclusión de sus comentarios (opción codificada como **Com**). Los textos de las respuestas para la opción “Comentarios” están disponibles en ANEXO 2, Textos correspondientes a las preguntas con opción a respuestas abiertas.

Solo dos de los encuestados aportaron información en esta modalidad: EN40 y EN42. El caso del servicio por el cual responde EN40 es una situación similar a la ya mencionada por R01 (respuesta recibida mediante correo-e) en el sentido de que es un servicio que produce diversos tipos de documentos y utilizan diferentes flujos de trabajo para cada tipo. En el caso de EN42, posiblemente utilicen *XML-out* (y así lo ha consignado como opción en la **Pregunta 5**) pero en su comentario aclara que no está del todo seguro.

## CAPÍTULO 8. CONCLUSIONES Y PROYECCIONES

De acuerdo con los objetivos propuestos para este trabajo y teniendo en cuenta los resultados obtenidos en el estudio exploratorio de la realidad actual de implantación JATS, se mencionan a continuación las conclusiones a las que se ha arribado. Las mismas se presentan en el orden en que fueron planteados los objetivos específicos.

De la descripción y el análisis realizado en el capítulo 3 sobre los cambios en las diferentes versiones y en el capítulo 4 sobre las líneas de diseño y las funcionalidades de JATS, se desprenden varias conclusiones relacionadas con la evolución del estándar, vinculadas al objetivo específico 1 de este trabajo. Se sintetizan las bases fundamentales del modelo para artículos, importantes para comprender como se plasmó el modelo en el diseño del esquema, así como también las razones de los cambios posteriores en el diseño que han marcado la evolución del estándar.

Las conclusiones relacionadas con el estado de implantación actual de JATS, vinculadas a los objetivos específicos 2 y 3 de este trabajo, son resultado del análisis y algunas reflexiones sobre los ejemplos seleccionados para el capítulo 6, que incluyen tanto implantación de JATS “tal cual es” como personalizaciones de JATS (abordadas en el capítulo 5).

Las conclusiones relacionadas con la situación actual en bibliotecas en cuanto a la implantación de JATS, planteado como objetivo 3 de este trabajo, se desprenden de los resultados del estudio de los ejemplos relacionados con bibliotecas abordados en el capítulo 6 y del análisis de la encuesta a bibliotecas incluido en capítulo 7.

### 8.1. Conclusiones

#### Objetivo 1. Estudiar las líneas principales que sustentan el diseño de JATS

##### **JATS ha evolucionado para atender a diversas tareas sobre los artículos: archivo, intercambio, publicación y autoría**

JATS es mucho más que una opción estandarizada para el archivo e intercambio de información, también es útil para la gestión de procesos editoriales, y para la creación de contenido. Tanto los proyectos editoriales en acceso abierto como los comerciales se benefician con la implantación del estándar JATS, pero en el caso del acceso abierto cobra más relevancia el hecho de aplicar estándares que tengan en cuenta la reutilización de contenidos.

Es justamente esa múltiple funcionalidad que caracteriza al estándar JATS la que promueve los cambios principales que ha tenido el modelo para artículos, evolucionado desde una única definición centrada en el archivo y el intercambio (1999) hasta la *Suite* y los tres modelos actuales, representados en los respectivos *Tag Sets* o conjuntos de etiquetas.

##### **Las sucesivas modificaciones en el diseño del esquema acompañan la evolución del modelo para artículos de PMC, incluyendo algunas modificaciones que han significado incompatibilidad con versiones anteriores**

En el caso de JATS, la evolución del esquema de origen creado por PMC en 1999 hacia otras versiones posteriores ha estado pautada por modificaciones tanto a nivel de componentes como de organización de la estructura del esquema que representa al modelo para artículos. Los cambios que se han implementado en las sucesivas versiones han sido motivados por diferentes razones, implicando modificaciones con distintas consecuencias a nivel de diseño.



No todas las modificaciones del esquema han implicado grandes cambios en el diseño. En las primeras versiones de NLM DTD los cambios son motivados por la necesidad de ampliar el marcado para incorporar cada vez más detalles del contenido de las revistas y más semántica, y se manifiestan con la adición de nuevos elementos. En los años siguientes se producen cambios estructurales más profundos originados por la necesidad de diferenciar el marcado según el cometido para el que se lo necesite. Eso se manifiesta en el diseño con la introducción de nuevos conjuntos de etiquetas, pasando de un único conjunto a dos y posteriormente se agrega un tercero, pues a los dos conjuntos existentes en NLM *Tag Suite* v.1.0 de 2003 (un conjunto orientado a la preservación y el intercambio y otro orientado a la publicación), se agrega en la v.2.0 de 2004 un tercer conjunto menos extenso y más estricto orientado a la autoría de artículos. Este cambio significó una versión con importantes diferencias respecto a la anterior, pues se realizó una remodelarización de la *Tag Suite* para satisfacer los nuevos requisitos.

Existen versiones en las que se realizaron solo cambios menores, para introducir mejoras solicitadas por NLM y otros usuarios, relacionados con los valores de algunos atributos. Otros cambios menores fueron originados en la necesidad de adaptación a las nuevas versiones del esquema MathML de W3C, debiéndose modificar los módulos del conjunto de etiquetas NLM para reflejar la nueva MathML.

Los cambios que significaron incompatibilidad con todas las versiones anteriores se dan en la versión 3.0 de NLM *Tag Suite*. Los cambios fueron muchos, algunos implicaron modificar el nombre de elementos, y hasta la adición de algunos elementos que pasan a cumplir la función de agrupar otros elementos ya existentes. En suma, hay cambios que han sido menores, otros que no han podido llevarse a cabo sin la remodelarización del conjunto de etiquetas y algunos incluso han implicado incompatibilidad con versiones anteriores del lenguaje. Los modelos que incluye el estándar actual son los mismos tres modelos base de NLM *Tag Suite* v.3.0, motivo por el cual JATS v.1.1 de 2015 es totalmente compatible con NLM versión 3.0.

### **JATS atiende a todos los campos científicos y cuenta con una comunidad de usuarios internacional**

Como se ya ha señalado en el capítulo 3, por sus características, NLM *Tag Suite* se convierte rápidamente en un lenguaje muy utilizado en el ámbito de las revistas del área ciencia, tecnología y medicina, trascendiendo ampliamente el ámbito del proyecto para el que fue creado. Actualmente JATS tiene un equipo y una comunidad de usuarios internacional. Está siendo utilizado en EEUU, Corea, Japón, Canadá y algunos países de Iberoamérica, África y Europa, por editoriales, y también archivos de revistas electrónicas en PubMed Central, Portico, la Biblioteca del Congreso de Estados Unidos y la Biblioteca Británica entre otros.

### **JATS atiende a todos los idiomas**

Cambios como los de la v.1.0 de JATS con la introducción de la codificación de idioma para documentos en los que algunos de los metadatos o el texto estén en varios idiomas, propiciaron la ampliación de los colectivos de usuarios que utilizan el estándar (como se comprueba en los portales de revistas de Japón y Corea). Ese cambio es una de las razones por las cuales revistas en idiomas con alfabetos no latinos marcan actualmente con JATS, con lo que el ámbito de utilización del estándar se amplía considerablemente.

### **JATS es un estándar abierto, de jure a nivel norteamericano y de facto en el resto del mundo**

Ser un estándar nacional garantiza a JATS el disponer del respaldo organizacional de NISO para sus actualizaciones y mantenimiento. El contar con un Comité Permanente y grupos de trabajo establecidos, además de un sistema para recoger las propuestas de cambios que hacen llegar los usuarios y ponerlas a consideración del Comité, asegura que JATS se mantiene actualizado y se nutre

de los aportes y las sugerencias de la comunidad. Todo ello da mayores garantías de continuidad a largo plazo. Es muy significativo el rol que desempeña NISO en la actualidad para JATS, coordinando a las comunidades de creadores de contenido, intermediarios y consumidores, continuando el trabajo que inició NLM, al convertir en estándar un modelo para artículos de revistas que ya era estándar *de facto*.

### **JATS resulta una buena opción para la preservación del contenido cuando las prácticas de marcado están documentadas y la calidad del XML asegurada**

Al ser JATS un estándar abierto, no propietario, bien documentado y con el respaldo de una agencia nacional de estandarización de Estados Unidos (NISO), se constituye en una buena opción para la preservación del contenido completo de las contribuciones a revistas electrónicas. Es utilizado por proyectos interesados en la preservación, aunque para su utilización se advierten algunos riesgos como la variabilidad en su implantación, la calidad del marcado o el problema de los artículos que tienen contenido externo asociado (datos y otros ficheros en diversos formatos relacionados con los artículos).

La variabilidad en las prácticas de marcado y la flexibilidad de los *Tag Sets* de Archivo e Intercambio y de Publicación, pueden presentar no solo los inconvenientes para la reutilización del contenido, sino que es además señalada como origen de uno de los principales riesgos para la preservación de ficheros JATS XML en la evaluación realizada por *British Library Preservation Team*.

Es necesario poner énfasis en la documentación de las prácticas existentes, con el objetivo de ayudar en el uso, y promover prácticas de marcado uniformes que aseguren la calidad. Cuando los comités de trabajo de NISO JATS estandarizan el lenguaje de marcado de NLM, también consideran importante documentar la forma de trabajar con él en la publicación de revistas, pues es fundamental para la reutilización e intercambio de contenido actual y a futuro de los artículos.

### **JATS es adaptable a necesidades particulares, permitiendo además crear nuevos conjuntos personalizados a partir de los proporcionados por el estándar**

La característica más relevante del estándar JATS en cuanto a diseño es el desarrollo de una *Suite* de definiciones que son reutilizables, como forma de contar con una definición de alcance amplio para el esquema. Esta *Suite* se acompaña de tres definiciones de propósito específico que han sido creadas para usos concretos. El diseño, en el caso de la codificación en DTD, está compuesto por JATS DTD *Suite* que ha sido escrita como una serie de módulos XML que se pueden combinar para ser utilizados en los diferentes *Tag Sets* del estándar, además de poderse utilizar para generar otros *Tag Sets* personalizados a partir del estándar (nuevos *Tag Sets*).

Como se ha mencionado en el capítulo 5, los conjuntos de etiquetas pueden ampliarse o restringirse para satisfacer las necesidades de un determinado proyecto, existiendo además la posibilidad de crear nuevos conjuntos de etiquetas a partir de los elementos y atributos de JATS. El diseño extensible es lo que ha hecho posible las personalizaciones actualmente existentes. Además, ha resultado muy provechoso el uso que algunos emprendimientos editoriales están propiciando en base a su implantación de personalizaciones de JATS, que facilitan la generación de otros productos como los ya mencionados en el capítulo 6: indicadores bibliométricos y cienciométricos (SciELO), fichas de especies y glosarios de términos taxonómicos (TaxPub), generación de tablas en índices (BITS).

La flexibilidad de los *Tag Sets* de Publicación, y de Archivo e Intercambio en particular, pensada para dar cabida a las necesidades de diferentes revistas, les ha permitido a las editoriales escoger qué elementos y atributos desean incluir. JATS es un esquema que admite la incorporación de semántica y taxonomías personalizadas, o ambos. Como consecuencia de ello, ha habido una considerable divergencia entre las diferentes editoriales en el marcado de los artículos. Esto ocasiona a veces que las equivalencias entre los contenidos para artículos de revistas de diferentes editoriales no sean

directas, lo cual genera inconvenientes pues aumenta la complejidad de la conversión, el intercambio de contenidos y la correcta lectura del contenido marcado por parte de los sistemas informáticos.

Como respuesta a esa situación, y con el objetivo de optimizar la reutilización de los contenidos de revista, es que aparecen grupos de trabajo como JATS4R, que tiene como cometido definir directrices sobre las mejores prácticas de marcado a modo de solucionar este inconveniente.

## **Objetivo 2. Investigar los usos específicos actuales que se sustenten en la aplicación de JATS**

### **El Tag Set elegido depende de las necesidades del usuario**

El conjunto de etiquetas o *Tag Set* que se escoge para implantación depende siempre de la funcionalidad que prima para satisfacer las necesidades de quien lo utiliza: archivar, convertir formatos, crear contenidos de revista, o publicar el contenido ya creado. Las formas de implantación estudiadas a través de los casos presentados en el capítulo 6 para ilustrar esos usos específicos actuales del estándar JATS así lo confirman.

### **El Tag Set de Archivo e Intercambio suele usarse con documentos originales en XML**

Los usuarios del *Tag Set* de Archivo e Intercambio son aquellos que ya tienen documentos XML y optan por convertidos a JATS XML, principalmente con la finalidad de archivarlos. Utilizan este *Tag Set* porque es el más flexible de los tres y por ello se requiere menos reorganización y reagrupamiento del contenido existente al convertirlo a JATS que si se empleara el *Tag Set* de Publicación.

JATS ha demostrado su valor como un método eficiente para el intercambio y la conversión de contenidos de revistas. La utilización de un estándar de marcado de amplia difusión y adopción facilita el intercambio de artículos entre diferentes editores, autores, agregadores de contenido, repositorios y servicios de indización. Utilizar JATS facilita la conversión de contenido y el intercambio de datos.

### **El Tag Set de de publicación, más prescriptivo, suele usarse con documentos originales no XML**

El conjunto de etiquetas de Publicación es utilizado mayoritariamente por aquellos que necesitan convertir contenido cuyo origen no es XML (ficheros que han sido generados con Word u otros formatos propietarios o no propietarios). Este conjunto les resulta muy adecuado pues permite etiquetar los metadatos que se originan durante el proceso de publicación, como la fecha de publicación, título, volumen y número de la revista, los números de página del artículo, etc. Ha resultado lo suficientemente flexible como para adaptarse a la mayoría de los estilos de las editoriales que lo implantan. Y cuando no, han surgido también personalizaciones de este *Tag Set*. Además, ha resultado muy útil para aquellos que al crear nuevos artículos marcan con este conjunto y luego utilizan el contenido JATS XML para la búsqueda y la visualización, o bien reutilizan fragmentos del contenido para otros propósitos (combinación en nuevos documentos, etc.).

### **El Tag Set de Autoría solo es utilizado en aplicaciones y sistemas que cuenten con herramientas de creación y edición en XML**

Los casos abordados en el capítulo 6 tratan de la implantación actual de JATS o sus personalizaciones relacionadas con los *Tag Sets* de Archivo e Intercambio y de Publicación. Esto es así porque si bien existe un tercer *Tag Set*, el de Autoría, en la práctica solo hay algunos escenarios actualmente donde éste puede ser utilizado. El motivo fundamental es que las herramientas de creación y edición de XML no son comunes en los programas de escritorio de propósito general. Lo utilizan mayormente aquellos autores que publican en editoriales que disponen actualmente de plataformas de gestión de contenido editorial en línea. Esas plataformas incluyen programas informáticos de gestión integral de la producción editorial y utilizan *XML-first* como forma implantar XML durante todo el proceso. Este *Tag Set* emplea el menor número de opciones de etiquetado

posibles para definir todo el contenido del artículo y no contempla metadatos que no se conocen al momento de la escritura del artículo, como la fecha de publicación, el título de la revista o las páginas que ocupará, etc.

**El Tag Set de Publicación es muy utilizado por las editoriales que optan por implantar el estándar “tal cual es”**

Las editoriales, tanto las que publican en acceso abierto, como los modelos híbridos o las comerciales, optan mayoritariamente por el Tag Set de Publicación, diseñado para proporcionar un marcado de contenidos que facilita la producción de artículos tanto para la web como impresos. El Tag Set de Publicación se presta para ello pues al ser más prescriptivo que el de Archivo e Intercambio el contenido es marcado de una manera más predecible, lo que facilita su tratamiento con herramientas informáticas.

**La implantación de JATS durante todo el proceso editorial se da mayormente relacionada con emprendimientos que tienen altos volúmenes de producción editorial**

La implantación varía, pues dependerá de qué tipo de flujo de trabajo con JATS XML sirva mejor su tamaño y modelo de negocio. Mientras que las editoriales pequeñas se vuelcan mayormente al *XML-out* (pues esta forma de trabajo les implica menores costos), las grandes editoriales invierten en la automatización de todo el proceso e implementan la forma *XML-in* en sus procesos editoriales. Para ello emplean plataformas de software (*Production Management Systems*) que se ocupan de toda la gestión del proceso editorial en XML. Lo que ambas formas de trabajo tienen en común es que utilizan un único fichero fuente en XML para generar sus productos finales.

En los casos estudiados se aprecia como los procesos editoriales se adaptan a la escala de cada negocio, y la inversión en tecnología para incorporar flujos de trabajo *XML-first* está mayormente relacionada con emprendimientos que tienen altos volúmenes de producción editorial. De todas formas, aunque el modelo de negocio editorial no sea de lucro, el factor económico siempre condiciona la sostenibilidad del proyecto.

**Las principales trabas para la adopción de XML en los procesos editoriales son: desconocimiento del proceso de producción de ficheros XML, no contar con personal capacitado y los costos asociados a la implantación**

Dentro de los obstáculos para la adopción de XML (que también afectan la adopción de JATS XML en particular) en los procesos editoriales se encuentra el desconocimiento del proceso de producción de ficheros XML por parte de los editores y demás personal relacionado con la publicación. No contar con personal capacitado es una traba para que muchos emprendimientos editoriales no implanten XML en sus procesos de producción. Los costos de implantación también pueden convertirse en otro obstáculo. Por otra parte, si no se cuenta con personal capacitado, pero hay recursos económicos suficientes, la alternativa existente es contratar empresas o particulares que brindan servicios editoriales.

**JATS beneficia a las editoriales medianas y pequeñas**

JATS beneficia a las editoriales medianas y pequeñas por varios motivos. Posibilita comenzar a crear artículos en XML significativamente más rápido, más fácil y con un coste menor, sin tener que crear un modelo y persuadir a los proveedores y socios de publicación para utilizarlo. Es considerablemente más fácil y menos costoso, sobre todo para las editoriales pequeñas, convertir archivos marcados con un estándar a otros formatos de visualización, pues ya existen XSLT y otras transformaciones listas para aplicar al archivo fuente JATS XML, posibilitando la obtención de ficheros en distintos formatos, compatibles con diferentes dispositivos. La posibilidad de generar y ofrecer

diferentes formatos de salida es importante para mejorar la competitividad en editoriales pequeñas y medianas, pues estarán brindando a sus clientes los formatos más actuales, convenientes para adaptar el despliegue del contenido a diversos dispositivos de lectura.

### **Objetivo 3. Aportar un panorama actual de la implantación de JATS en bibliotecas**

**En las bibliotecas prima el uso del *Tag Set* de Archivo e Intercambio. Es utilizado para archivar, convertir desde otros formatos y para reutilizar la información en la creación de otros productos**

En el caso de las bibliotecas, especialmente las nacionales (LC, BL), prima el uso de JATS XML como formato para el archivo, utilizan el *Tag Set* de Archivo e Intercambio y lo emplean para la conversión desde otros formatos que reciben. Emplean JATS XML para sus archivos digitales de *e-journals*, los cuales utilizan el estándar JATS en sus procesos internos para normalizar el contenido convirtiendo desde diversos formatos a JATS para unificar contenidos (PubMed, Portico, Scholars Portal). Utilizan el *Tag Set* de Archivo e Intercambio porque es el que mejor se desempeña cuando se trata de respetar la mayor cantidad posible de semántica en el contenido XML que provenga de otros *Tag Sets*.

Que la LC ponga a JATS como primera opción para el depósito de revistas en su programa de depósito legal para publicaciones seriadas es significativo, pues las decisiones de esta biblioteca han sido históricamente de gran influencia en el mundo bibliotecario en lo que respecta a las reglas, estándares y formatos bibliográficos que adopta.

### **La mayoría de bibliotecas académicas reconocen las ventajas de XML**

Las bibliotecas reutilizan la información para la creación de otros productos aprovechando la información marcada con JATS XML (en catálogos, bases de datos, servicios de indización y resúmenes), extrayendo a partir del fichero XML la información específica que se necesite de forma automatizada, para generar otros productos de información diferentes a partir de ese fichero fuente.

La mayor parte de las bibliotecas que han respondido la encuesta reconocen varias de las ventajas de utilizar XML y las relacionan con el trabajo bibliotecario: reutilizar el contenido en catálogos, en sitios web, en bases de datos, generar productos impresos, automatizar los procesos, acompañar el crecimiento de los servicios y preservar el contenido digital para el futuro. Las principales ventajas para implantar XML en la publicación de contenidos según los encuestados, mencionadas en orden de importancia decreciente son: disponibilizar contenido rápida y fácilmente en varios formatos electrónicos (ej. ePub, HTML, PDF), preservar el contenido digital para el futuro, reutilizar el contenido en otros canales (catálogos, sitios web, bases de datos, impresos) y automatizar procesos.

### **Las bibliotecas académicas han comenzado a utilizar XML para prestar servicios editoriales**

En cuanto a las bibliotecas, utilizan XML para varios propósitos. El interés en utilizar XML por parte de las bibliotecas en un principio estuvo sustentado principalmente en la preocupación por preservar y recuperar a largo plazo la información contenida en las revistas electrónicas. A los archivos les interesa estandarizar para cumplir su cometido, formando parte de su estrategia de preservación. Más recientemente, también se da la tendencia de utilizar XML como tecnología aplicada a la prestación de servicios editoriales, principalmente en las bibliotecas de instituciones académicas y de investigación, más aún en aquellas asociadas a las editoriales universitarias.

**El modelo de prestación de servicios editoriales en las bibliotecas académicas tiene las características propias de una actividad a pequeña escala y no necesariamente implica la utilización de XML**



El modelo de prestación de servicios editoriales de las bibliotecas es una actividad de pequeña escala, sin fines de lucro, generalmente orientada a la publicación en acceso abierto y basada en el presupuesto de la biblioteca para solventar sus actividades. Sin embargo, las bibliotecas que brindan estos servicios pasan a llenar una necesidad importante proporcionando apoyo y visibilidad a las revistas de su institución.

El rol que cumpla la biblioteca respecto a la publicación también determina si necesita utilizar XML o no. Muchas de las bibliotecas que respondieron a la encuesta solamente mantienen un archivo de publicaciones y dan acceso a las mismas (hosting, repositorios). Solo unas pocas (asociadas o no a la editorial universitaria) brindan servicios editoriales profesionales que implican de alguna manera componer, editar, generar servicios en base al contenido. Algunas complementan además lo anteriormente mencionado con brindar asesoramiento en servicios editoriales.

Existen además diferentes niveles de servicio relacionados con la forma en que las bibliotecas proporcionan servicios de publicación orientados a revistas, dependiendo del involucramiento de la biblioteca y los roles ésta que asuma. La mayoría de ellas ofrecen un sistema en línea y el correspondiente soporte técnico, mientras que otras actividades relacionados con la edición de artículos, correcciones de estilo, etc., siguen siendo responsabilidad del editor de cada revista. Si bien la utilización de XML puede significar una serie de ventajas para las bibliotecas que prestan servicios editoriales o planean hacerlo en el futuro, en la realidad actual de las bibliotecas objetivo de la encuesta, solo algunas se involucran a fondo con la implantación de flujos de trabajo XML.

Para algunas bibliotecas, la presentación del contenido de las revistas en la Web como archivos PDF descargables que incluyan capacidad de búsqueda mediante OCR puede ser suficiente. Ahora bien, si planean ofrecer contenidos en diferentes formatos (PDF, ePub, mobi, HTML, etc.), reutilizar contenidos para otros propósitos (catálogos, sitios web, bases de datos, etc.), asegurar que el contenido pueda ser migrado a nuevos formatos en el futuro, entonces es muy recomendable que comiencen a utilizar XML.

**Las bibliotecas que no usan XML explican sus motivos: opción inexistente en el sistema, escaso volumen de trabajo editorial, falta de personal capacitado y de adaptación de los flujos XML a los productos concretos**

El amplio porcentaje de las bibliotecas que respondieron la encuesta brindan servicios de publicación sin implementar XML. La razón que aparece en la encuesta en primer lugar es que el sistema de publicación que utilizan actualmente no incluye la opción XML. Las bibliotecas están cómodas con sus sistemas actuales ya que otra de las razones más importantes que se detecta en la encuesta es que los editores de las revistas que ellos publican no requieren específicamente XML. Por otra parte, algunas bibliotecas no consideran necesario implantar XML debido al poco volumen de trabajo editorial que manejan actualmente.

La tercera razón de peso por la cual las bibliotecas encuestadas no incorporan XML actualmente es que no cuentan con personal capacitado. Esa carencia puede muy bien constituirse en un nicho de trabajo para los bibliotecarios con conocimientos en el área de la edición digital. En las bibliotecas que figuran en el LPD 2016 las actividades relacionadas con publicación digital tienen un espacio propio dentro del organigrama de la biblioteca. Se menciona en muchos casos la existencia de una sección o bien de un encargado de publicación digital. Las actividades editoriales de la biblioteca se encuentran insertas en diferentes secciones de la biblioteca, siendo asumidas por bibliotecarios en diferentes niveles de carrera.



Otras razones vinculadas a la no adopción de XML por parte de las bibliotecas objeto de la encuesta es que los flujos XML no se adaptan a los productos que publican. Los flujos de trabajo XML en la producción editorial están orientados actualmente a la publicación de contenidos académicos tradicionales, a la reutilización de contenido y a la preservación de contenidos a largo plazo, pero no se adaptan muy bien a otras formas de publicación académica no tradicionales. A medida que la publicación de contenidos académicos en la *web* incorpora múltiples tipos de contenido, interacción y multimedia, los flujos de trabajo basados en XML pueden no siempre ser la manera más eficiente de prestar servicios editoriales orientados a esos tipos de productos.

### **Ciertas habilidades y competencias incluidas en la formación bibliotecaria resultan muy útiles en el ámbito editorial**

Los bibliotecarios tienen habilidades y competencias profesionales que pueden aportar al ámbito editorial, como las habilidades tecnológicas incluidas en la formación bibliotecaria: cursos en lenguajes de marcado XML, PHP y SQL para aplicaciones de bases de datos, etc., cada vez más importantes para publicar en digital. Otra ventaja de la formación bibliotecaria para el ámbito editorial son los conocimientos sobre preservación digital, sumamente útiles en un contexto en el cual los documentos originales están naciendo digitalmente. Si bien la preservación no forma parte del proceso de publicación, los metadatos que se generan en el proceso sí, y la creación y gestión de metadatos son una competencia habitual en los bibliotecarios.

### **Son pocas las bibliotecas académicas que implantan XML para producción editorial, y dentro de éstas solo algunas están utilizando actualmente o piensan utilizar JATS en el futuro**

Menos del 30% de las bibliotecas que respondieron la encuesta utilizan o estar pensando utilizar XML para LPS. Entre las que afirmaron que implantan XML, luego de contrastar las respuestas a diferentes preguntas en análisis de la encuesta, ha resultado que solo once bibliotecas lo implantan en procesos editoriales. Las demás lo utilizan para otros fines, mayormente relacionados con conversión de formatos y con la reutilización del contenido marcado con XML en otros productos que ofrecen.

El porcentaje de bibliotecas que efectivamente utilizan JATS XML en sus procesos editoriales actualmente o están pensando en implantarlo en un futuro representa menos de un 2% del total de las bibliotecas que han respondido a la encuesta. El resto de las bibliotecas que utilizan XML implantan TEI en su gran mayoría. Esto puede deberse al hecho que TEI está más difundido, pues hace más de 20 años que existe ese lenguaje. JATS en cambio, es mucho más reciente como opción estandarizada para el marcado.

## **8.2. Proyecciones**

El carácter exploratorio de este trabajo ha propiciado la generación de posibles nuevos temas relacionados con estándares y tecnologías de marcado. La temática estudiada no se agota aquí, sino que pretende ser punto de partida para otros estudios que puedan tomar los resultados o partes del desarrollo de este trabajo como antecedentes, utilizando lo que resulte adecuado según sus vinculaciones a futuras propuestas que se puedan desarrollar.

Un importante aporte de este trabajo es satisfacer la necesidad de contar con información actual sobre la implantación de JATS. Tanto para nuevos emprendimientos editoriales como para aquellos existentes que quieran cambiar su modo de trabajo actual y comenzar a implantar XML, podrá

servir como referencia para el desarrollo de proyectos propios, aportando información sobre la opción JATS como tecnología de marcado.

Otro aspecto que aporta valiosa información para futuras investigaciones es el detalle de los antecedentes, cambios en el diseño y estructura modular del estándar, que pueden resultar muy útiles en el caso de futuros estudios que se plantearan comparar JATS con otras opciones de marcado para artículos de revistas.

Este trabajo deja además algunos temas planteados vinculados a la incursión de las bibliotecas en la prestación de servicios editoriales, como la inquietud por la especialización del perfil bibliotecario para integrar los equipos de trabajo que brindan este tipo de servicios.

En el terreno de las habilidades, destrezas y competencias que incluye la formación bibliotecaria para la producción de contenidos digitales, algunos autores a los que se hace referencia en este trabajo son de la opinión de que la formación actual en esta área no está consolidada. Sería interesante explorar ese aspecto teniendo en cuenta las mallas curriculares de diversas Universidades, institutos terciarios, etc.

## REFERENCIAS

- Abrams, S. L., & Rosenblum, B. (2003). XML for e-journal archiving. *OCLC Systems & Services*, 19(4), 155–161.
- Bazargan, K. (2015). A complete end-to-end publishing system based on JATS. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2015*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK279828/>
- Beagrie, C. (2013). *Preservation, Trust and Continuing Access for e-Journals*. Digital Preservation Coalition. Retrieved from [http://www.dpconline.org/component/docman/doc\\_download/924-dpctw13-04](http://www.dpconline.org/component/docman/doc_download/924-dpctw13-04)
- Beck, J. (2010). Report from the Field: PubMed Central, an XML-based Archive of Life Sciences Journal Articles. In *Proceedings of the International Symposium on XML for the Long Haul: Issues in the Long-term Preservation of XML. Balisage Series on Markup Technologies*, vol. 6. doi:10.4242/BalisageVol6.Beck01
- Beck, J. (2011). NISO Z39.96 The Journal Article Tag Suite (JATS): What Happened to the NLM DTDs? *The Journal of Electronic Publishing : JEP*, 14(1). doi:10.3998/3336451.0014.106
- Beck, J. (2013). What JATS Users should Know about the Book Interchange Tag Suite (BITS). In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2013*. Bethesda (MD): National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK159737/>
- Beck, J., Chodacki, J., Eaton, A., Evans, M., Gatti, R., Gilbert, J., ... Mowlam, T. (2015). Improving the reusability of JATS. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK279901/>
- Beck, J., & Randall, L. (2013). NLM DTD to NISO JATS Z39.96-2012. In *The NCBI Handbook [Internet]* (2nd ed.). National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK169004/>
- Beck, J., & Usdien Tommie, B. (2013). JATS - A new standard for an old specification. *Information Standards Quarterly*, 25(19-21). Retrieved from [http://www.niso.org/apps/group\\_public/download.php/10418/SP\\_Beck\\_Usdin\\_JATS\\_isqv25no1.pdf](http://www.niso.org/apps/group_public/download.php/10418/SP_Beck_Usdin_JATS_isqv25no1.pdf)
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. (2008). Extensible Markup Language ( XML ) 1.0 ( Fifth Edition ). Retrieved from <http://www.w3.org/TR/2008/REC-xml-20081126/>
- Brown, A. (2003). XML in serial publishing: Past, present and future. *OCLC Systems and Services*, 19(4), 149–154. Retrieved from <http://search.proquest.com/docview/209776679?accountid=14501>
- Carpenter, T. (2014). JATS and NISO — the value of community standardization. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2013/2014 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK195186/>
- Cave, F. (2003). Article metadata standards: an historical review. *OCLC Systems and Services*, 19(4), 144–148. Retrieved from <http://search.proquest.com/docview/209773731?accountid=14501>
- Delfino, E., & Mandelbaum, J. (2011). eDeposit for eSerials: Current Work and Plans at the Library of Congress. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK62174/>

- Digital Preservation Coalition. (2015). Digital Preservation Handbook, 2nd Edition. Retrieved from <http://www.dpconline.org/advice/preservationhandbook>
- Directory of Electronic Journals, Newsletters, and Academic Discussion Lists, Volumes 1-2.* (1991). Washington DC: Association of Research Libraries. Retrieved from <https://books.google.com/books?id=KmrGAAAAMAAJ&pgis=1>
- Donohoe, P., Sherman, J., & Mistry, A. (2015). The Long Road to JATS. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2015*. Bethesda (MD): National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK279831/>
- Eden, M., & Cleghorn, T. (2016). An Implementation of BITS: The Cambridge University Press Experience. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2016 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK350535/>
- Fadik, B., Gross, M., & Trombley, B. (2015). Automating Complex High-Volume Technical Paper and Journal Article Page Composition with NLM XML and InDesign. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK279830/>
- Fonseca, G. (2016). Pergunta sobre a especificação SciELO PS - Google Groups. *SciELO XML - SciELO Publishing Schema*. Retrieved March 29, 2016, from <https://groups.google.com/forum/#!topic/scielo-xml/69zidabhimg>
- Graham, T. (2014). Formatting JATS: as easy as 1-2-3. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2013/2014*. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK189779/>
- Gross, M. (2003). Data capture & conversion. In W. E. Kasdorf (Ed.), *Columbia guide to digital publishing* (pp. 179–218). New York: Columbia University Press.
- Gutzman, K., & Tryka, K. A. (2013). Extending JATS to include the NISO/NFAIS Recommended Practices for Online Supplemental Journal Article Materials. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK159734/>
- Hahn, K. L. (2008). *Research Library Publishing Services: new options for University publishing*. Washington DC: ARL. Retrieved from <http://www.arl.org/storage/documents/publications/research-library-publishing-services-mar08.pdf>
- Harrison, M. (2014). Production as a specialist skill: the eLife method. In *E-Production Seminar 2014*. International Association of STM Publishers. Retrieved from <http://www.stm-assoc.org/events/e-production-seminar-2014/?presentations>
- Harrison, M. (2016). Collecting XML at article submission at eLife: two steps forward, one step back? In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2016 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK350147/>
- Hawkins, K. S. (2015). The Evolution of Publishing Agreements at the University of Michigan Library. In M. Bonn & M. Furlough (Eds.), *Getting the Word Out: Academic Libraries as Scholarly Publishers* (pp. 113–118). Chicago, Illinois: ACRL. Retrieved from [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838986981\\_getting\\_OA.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838986981_getting_OA.pdf)
- Huh, S. (2013). ScienceCentral: open access full-text archive of scientific journals based on Journal Article Tag Suite regardless of their languages. *Biochemia Medica*, 23(3), 235–6. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3900078&tool=pmcentrez&render>

type=abstract

- Huh, S. (2014). Coding practice of the Journal Article Tag Suite extensible markup language. *Science Editing*, 1(2), 105–112. doi:2014.1.2.105
- Huh, S., Choi, T. J., & Kim, S.-H. (2014). Using Journal Article Tag Suite extensible markup language for scholarly journal articles written in Korean. *Science Editing*, 1(1), 19–23. doi:10.6087/kcse.2014.1.19
- Huwe, T. K. (2014). Next-Generation Library Publishing: It's Here. *Computers in Libraries*, 34(8), 31–33. Retrieved from <http://search.proquest.com/docview/1625138508?accountid=14501>
- Hyde, K. W. (2015). *The library needs to determine an eDeposit and eCollections strategy*. Audit Report No. 2014-PA-101. Washington DC. Retrieved from <https://www.loc.gov/portals/static/about/documents/edeposit-and-ecollections-strategy-april-2015.pdf>
- Imsieke, G. (2013). A Publisher's InDesign to BITS and EPUB Infrastructure: Conventions, Configuration, Conversion, Checks. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK159733/>
- Inera. (2001). *E-journal archive DTD feasibility study*. Retrieved from <http://old.diglib.org/preserve/hadtdfs.pdf>
- ISO. ISO 12083:1994 - Information and documentation -- Electronic manuscript preparation and markup (1994). Retrieved from [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=20866](http://www.iso.org/iso/catalogue_detail.htm?csnumber=20866)
- JATS: Journal Article Tag Suite, version 1.1. (2015). Baltimore, MD: NISO. Retrieved from [http://www.niso.org/apps/group\\_public/download.php/15933/z39\\_96-2015.pdf](http://www.niso.org/apps/group_public/download.php/15933/z39_96-2015.pdf)
- Kasdorf, W. E. (2003). *The columbia guide to digital publishing*. New York: Columbia University Press.
- Keener, A. (2014). Library (Publishing) School: Training and Competencies for the New Publishing Professional. *The Journal of Electronic Publishing*, 17(2). doi:10.3998/3336451.0017.206
- Keishi, K., Tokushige, K., & Mitsuru, K. (2012). Reducing costs and expanding XML submissions with PDF to JATS conversion. In National Center for Biotechnology Information (US) (Ed.), *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2012* (pp. 1–7). Bethesda (MD). Retrieved from [http://www.ncbi.nlm.nih.gov/books/NBK100490/pdf/Bookshelf\\_NBK100490.pdf](http://www.ncbi.nlm.nih.gov/books/NBK100490/pdf/Bookshelf_NBK100490.pdf)
- Kenney, A. R., Entlich, R., Hirtle, P. B., McGovern, N. Y., & Buckley, E. L. (2006). *E-Journal Archiving Metes and Bounds: A Survey of the Landscape*. Washington DC: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub138/pub138.pdf>
- Kimura, H., & de Araujo Machuca, N. C. (2014). O Formato XML SciELO na RAC. *RAC - Revista de Administração Contemporânea*, 18(5), 1–3. doi:10.1590/1982-7849rac20140029
- Kraetke, M., & Buehring, F. (2015). A Quality Assurance Tool for JATS/BITS with Schematron and HTML reporting. *JATS-Con 2016 [Abstracts]*. Retrieved from <http://jats.nlm.nih.gov/jats-con/2016/schedule2016a.html>
- Lach Arlinghaus, S., & Zander, R. H. (2008). Electronic journals: then and now... a fifteen year retrospective. *Solstice*, 19(2). Retrieved from <http://www.mobot.org/plantscience/ResBot/Repr/Arling-Zand-ElecJour-Solstice08c31.pdf>
- Lancaster, F. W. (1995). The evolution of electronic publishing. *Library Trends*, 43(4), 518–27. Retrieved from [https://www.ideals.illinois.edu/bitstream/handle/2142/7981/librarytrendsv43i4c\\_opt.pdf](https://www.ideals.illinois.edu/bitstream/handle/2142/7981/librarytrendsv43i4c_opt.pdf)
- Lapeyre, D. A. (2008). A non-backwards-compatible update: a difficult decision. In *International Symposium on Versioning XML Vocabularies and Systems*. *Balisage Series on Markup*





- Technologies*, vol. 2. Montréal, Canada. doi:10.4242/BalisageVol2.Lapeyre01
- Lapeyre, D. A. (2009). The National Library of Medicine Tag Suite for Journal Articles: Taking Over the World of XML Journal Publishing. In *XML-in-Practice 2009* (p. 22 p.). Washington DC. Retrieved from [http://www.mulberrytech.com/papers/Lapeyre\\_gov.zip](http://www.mulberrytech.com/papers/Lapeyre_gov.zip)
- Lapeyre, D. A., & Usdin, B. T. (2011). Introduction to Multi-language Documents in NISO JATS. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK62175/>
- Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE*, 10(6), e0127502. Retrieved from <http://dx.doi.org/10.1371/journal.pone.0127502>
- Library of Congress. (2006). LOC, British Library Support Standard for Electronic Journals - News Releases. *News from the Library of Congress*. 2006 April 19. Retrieved April 18, 2016, from <http://www.loc.gov/today/pr/2006/06-097.html>
- Lippincott, S. K. (Ed.). (2014). *Library Publishing Directory 2014*. Atlanta, GA: Library Publishing Coalition.
- Lippincott, S. K. (Ed.). (2015a). *Library Publishing Directory 2015*. Atlanta, GA: Library Publishing Coalition. Retrieved from [http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/lpc\\_dir\\_2015lpcd.pdf](http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/lpc_dir_2015lpcd.pdf)
- Lippincott, S. K. (Ed.). (2015b). *Library Publishing Directory 2016*. Atlanta, GA: Library Publishing Coalition. Retrieved from [http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/Library\\_Publishing\\_Directory\\_2016.pdf](http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/Library_Publishing_Directory_2016.pdf)
- Lorimer, R. (2013). Digital Developments in Libraries, Journals, and Monograph Publishing: Emerging Pitfalls, Practices, and Possibilities. *Scholarly and Research Communication*. Retrieved from <http://src-online.ca/index.php/src/article/view/43/118>
- Maloney, C. (2011). JATSPack and JATSPAN, a packaging format and infrastructure for the NLM/NISO Journal Archiving Tag Suite (JATS). *Balisage Series on Markup Technologies*, 7. doi:10.4242/BalisageVol7.Maloney01
- Mattson, M., & Friend, L. (2014). A planning perspective for library journal publishing services. *OCLC Systems & Services: International Digital Libraries Perspectives*, 30(3), 178–191.
- Maxwell, J. W., MacDonald, M., Nicholson, T., Halpape, J., Taggart, S., & Binder, H. (2010). XML Production Workflows? Start with the Web. *The Journal of Electronic Publishing*, 13(1). doi:10.3998/3336451.0013.106
- Mcglone, J. (2013). Preserving and Publishing Digital Content Using XML Workflows. In A. P. Brown (Ed.), *The Library Publishing Toolkit* (pp. 97–108). Geneseo, NY: IDS Project Press. Retrieved from <http://hdl.handle.net/2027.42/99563>
- Mietchen, D., McEntyre, J., Beck, J., Maloney, C., & Group, F. D. C. I. (2015). Adapting JATS to support data citation. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK280240/>
- Missingham, R., & Kanellopoulos, L. (2014). University presses in libraries: potential for successful marriages. *OCLC Systems & Services: International Digital Library Perspectives*, 30(3), 158–166. doi:10.1108/OCLC-01-2014-0001
- Momen, H. (2014). Megarevistas y otras innovaciones en la edición de revistas académicas [video]. In





- SciELO 15 Years Conference*. World Health Organization. Retrieved from <http://www.scielo15.org/es/video-megarevistas-y-otras-innovaciones-en-la-edicion-de-revistas-academicas/>
- Morrissey, S. M. (2011). "More What You'd Call "Guidelines" Than Actual Rules' : Variation in the Use of Standards. *The Journal of Electronic Publishing*, 14(1). doi:10.3998/3336451.0014.104
- Morrissey, S. M., Meyer, J., Bhattarai, S., Kurdikar, S., Ling, J., Stoeffler, M., & Thanneeru, U. (2010). Portico: A Case Study in the Use of the Journal Archiving and Interchange Tag Set for the Long Term Preservation of Scholarly Journals. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK47087/>
- Murray-Rust, P., & Rzepa, H. S. (2003). XML for scientific publishing. *OCLC Systems and Services*, 19(4), 162–169. Retrieved from <http://search.proquest.com/docview/209770934?accountid=14501>
- National Information Standards Organization. (2012). ANSI/NISO Z39.96-2012 JATS: Journal Article Tag Suite -. Retrieved from [http://www.niso.org/apps/group\\_public/project/details.php?project\\_id=93](http://www.niso.org/apps/group_public/project/details.php?project_id=93)
- NISO. (2015). NISO Launches New Project to Develop XML Standard for Producing Standards Documents - National Information Standards Organization. Retrieved October 11, 2015, from [http://www.niso.org/news/pr/view?item\\_key=02f81eb137174fc8c436264b55d6ffe13a8887b2](http://www.niso.org/news/pr/view?item_key=02f81eb137174fc8c436264b55d6ffe13a8887b2)
- O'Brien, D. (2011). Why Create a Customization of a Standard? An ACS Case Study. *The Journal of Electronic Publishing*, 14(1). doi:10.3998/3336451.0014.105
- O'Brien, D., Fisher, J., & Haines, D. J. (2011). Why Create a Customization of a Standard? An ACS Case Study. *The Journal of Electronic Publishing*, 14(1). doi:10.3998/3336451.0014.105
- O'Connor, C., Haenel, S., Gnanapiragasam, A., Hepp, M., & Fleischer, T. (2015). Building an Automated XML-Based Journal Production Workflow. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2015*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK279927/>
- Oakes, G. (2014). Automated page composition. In *E-Production Seminar 2014*. International Association of STM Publishers. Retrieved from [http://www.stm-assoc.org/2014\\_12\\_04\\_E\\_Production\\_Seminar\\_Oakes\\_Automated\\_page\\_composition.pdf](http://www.stm-assoc.org/2014_12_04_E_Production_Seminar_Oakes_Automated_page_composition.pdf)
- Owens, E. (2010). The Evolving Information Ecostructure of Publishing. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings*. Bethesda (MD): National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK48528/>
- Packer, A., Cop, N., Luccisano, A., Ramalho, A., & Spinak, E. (2014). *SciELO - 15 Years of Open Access: an analytic study of Open Access and scholarly communication*. Paris: UNESCO. doi:10.7476/9789230012373
- Padua, J., Ahaman, I., Apeztecha, H., & Borsotti, C. (1996). *Técnicas de investigación aplicadas a las ciencias sociales*. México: FCE.
- Penev, L., Catapano, T., Agosti, D., Georgiev, T., Sautter, G., & Stoev, P. (2012). Implementation of TaxPub, an NLM DTD extension for domain-specific markup in taxonomy, from the experience of a biodiversity publisher. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK100351/>
- Perera, C. (2016). Challenges in Implementing a Multilingual JATS Publishing Workflow. In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2016 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK355223/>

- Publishers Press. (2016). How can a publisher integrate an XML workflow? Retrieved January 14, 2016, from <http://www.pubpress.com/faq/how-can-publisher-integrate-xml-workflow>
- Rech, D. A. (2012). Instituting an XML-first workflow. *Publishing Research Quarterly*, 28(July), 192–196. doi:10.1007/s12109-012-9278
- Rosenblum, B. D. (2015). Smoke Screens of Impossibility. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK285911/>
- Rosenblum, B., & Golfman, I. (2002). A Decade of DTDs and SGML in Scholarly Publishing: What Have We learned? In *Extreme Markup Languages 4-9 agosto 2002*. Montréal. Retrieved from <http://conferences.idealliance.org/extreme/html/2002/Rosenblum01/EML2002Rosenblum01.html>
- Salminen, A., Jauhiainen, E., & Nurmeksela, R. (2014). A life cycle model of XML documents. *Journal of the Association for Information Science & Technology*, 65(12), 2564–2580. Retrieved from 10.1002/asi.23148
- Schwarzman, A. B., Hur, H., Pai, S.-L., & Glass, C. M. (2004). XML-centric workflow offers benefits to scholarly publishers. In *XML 2004 Proceedings*. Retrieved from <http://people.ischool.berkeley.edu/~glushko/202/Schwarzman-XML-CentricWorkflow.pdf>
- Sharp, M. (2013). Structured Documents for Science: JATS XML as Canonical Content Format | PLOS Tech. Retrieved April 13, 2015, from <http://blogs.plos.org/tech/structured-documents-for-science-jats-xml-as-canonical-content-format/>
- Skinner, K., Lippincott, S., Speer, J., & Walters, T. (2014). Library-as-Publisher: Capacity Building for the Library Publishing Subfield. *The Journal of Electronic Publishing*, 17(2). doi:10.3998/3336451.0017.207
- Spiro, L. (2015). Nimble and Oriented towards Teaching and Learning Publishing Services at Small Academic Libraries. In M. Bonn & M. Furlough (Eds.), *Getting the Word Out: Academic Libraries as Scholarly Publishers* (pp. 141–174). Chicago: ACRL. Retrieved from [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838986981\\_getting\\_OA.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/9780838986981_getting_OA.pdf)
- Stoker, T., & Rose, K. (2012). ACS Publications — Ensuring XML Quality. Presented at International Symposium on Quality Assurance and Quality Control in XML, Montréal, Canada, August 6, 2012. In *Proceedings of the International Symposium on Quality Assurance and Quality Control in XML. Balisage Series on Markup Technologies, vol. 9 (2012)*. doi:10.4242/BalisageVol9.Rose01
- Usdin, B. T. (2016). So You Want to Adopt JATS. What Decisions Do You Need To Make? In *Journal Article Tag Suite Conference (JATS-Con) Proceedings 2016 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK350380/>
- W3C. (2000). XML DTD Modules. Retrieved July 4, 2015, from [http://www.w3.org/TR/xhtml1-building/xml\\_dtd.html](http://www.w3.org/TR/xhtml1-building/xml_dtd.html)
- Ware, M., & Mabe, M. (2015). *The STM report: an overview of scientific and scholarly journal publishing*. The Hague. Retrieved from [http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf)
- Wheatley, P., May, P., Day, M., Kirchoff, A., & Meyer, J. (2015). *Preservation Assessment: JATS/NLM Format Preservation Assessment*. (British Library Digital Preservation Team, Ed.). London: British Library. Retrieved from [http://wiki.dpconline.org/images/2/21/JATS\\_NLM\\_Assessment\\_v1.3.pdf](http://wiki.dpconline.org/images/2/21/JATS_NLM_Assessment_v1.3.pdf)
- Wheeler, R., Rosenblum, B., & West, L. (2016). NISO STS Project Overview and Update. In *Journal*



- Article Tag Suite Conference (JATS-Con) Proceedings 2016 [Internet]*. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK350146/>
- Wittenberg, K., Kirchoff, A., & Morrissey, S. (2014). Scaling Up and Scaling Out: Leveraging Preservation Infrastructure and Experience to Benefit the Community. *Portico*. Retrieved from [http://www.portico.org/digital-preservation/wp-content/uploads/2015/05/Wittenberg\\_ScalingUpScalingOut.pdf](http://www.portico.org/digital-preservation/wp-content/uploads/2015/05/Wittenberg_ScalingUpScalingOut.pdf)
- Wusteman, J. (2003a). XML and e-journals. *OCLC Systems and Services*, 19(4), 125–131. Retrieved from <http://search.proquest.com/docview/209779628?accountid=14501>
- Wusteman, J. (2003b). XML and e-journals: The state of play. *Library Hi Tech*, 21(1), 21–33. Retrieved from <http://search.proquest.com/docview/200617553?accountid=14501>
- Zhao, W., David, R. H., Khwaja, S., & Lin, Q. (2015). JATS for E-journals and BITS for Ebooks---Adopting BITS for Scholars Portal Ebook Repository. National Center for Biotechnology Information (US). Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK280069/>

## LISTADO DE SITIOS WEB

Add a journal to PMC <http://www.ncbi.nlm.nih.gov/pmc/pub/pubinfo/>

ACS Publications <http://pubs.acs.org/>

Author Guide y Policies [eLife] [http://submit.elifesciences.org/html/elifelife\\_author\\_instructions.html](http://submit.elifesciences.org/html/elifelife_author_instructions.html)

BITS: schemas (DTD, XSD, RNG) and documentation

<ftp://ftp.ncbi.nlm.nih.gov/pub/jats/extensions/bits/1.0/>

Book Interchange Tag Suite (BITS) version 1.0 <http://jats.nlm.nih.gov/extensions/bits/#id48297>

Book Interchange Tag Set: JATS Extension <http://jats.nlm.nih.gov/extensions/bits/>

Certification Criteria for XML Article text Tagging Services according to SciELO Publishing Schema

<http://www.scielo.org/php/level.php?lang=enycomponent=42yitem=32>

Como produzir arquivo XML usando o Markup [http://scielo-pc-programs.readthedocs.org/en/latest/pt\\_how\\_to\\_generate\\_xml.html](http://scielo-pc-programs.readthedocs.org/en/latest/pt_how_to_generate_xml.html)

Document Hierarchy Diagrams. Journal Publishing Tag Library NISO JATS Draft Version 1.1d3 Version of April 2015 <http://jats.nlm.nih.gov/publishing/tag-library/1.1d3/chapter/nfd.html>

DTD ISO Standards Tag Set (ISOSTS) v1.1 20130701

<http://www.iso.org/schema/isosts/v1.1/ISOSTS.dtd>

eLife <http://elifesciences.org/>

Hogrefe Book Tag Set <https://hobots.hogrefe.com/>

How to make new tag sets <http://jats.nlm.nih.gov/publishing/tag-library/1.0/?chap=pub-implementor>

How to submit a file after validating the JATS XML file <http://e-sciencecentral.org/pub/filespec/>

International Association of STM Publishers. <http://www.stm-assoc.org/>

Introduction to Parameter Entities <http://jats.nlm.nih.gov/archiving/tag-library/1.1d1/n-hhx2.html>

ISOSTS <http://www.iso.org/schema/isosts/>

J-STAGE. Japan Science and Technology Agency <https://www.jstage.jst.go.jp/>

JATS. Journal Article Tag Suite. <http://jats.nlm.nih.gov/>

JATS: Versions. <http://jats.nlm.nih.gov/versions.html>

JATS Archiving DTD <ftp://ftp.ncbi.nlm.nih.gov/pub/jats/archiving/1.0/>

JATS-Con Home <http://jats.nlm.nih.gov/jats-con/>

JATS-Con Proceedings <http://www.ncbi.nlm.nih.gov/books/NBK65129/>

JATS-List. Open Forum on the Journal Article Tag Suite <http://www.mulberrytech.com/JATS/JATS-List/>

Journal Publishing Tag Library NISO JATS Version 1.0 <http://jats.nlm.nih.gov/publishing/tag-library/1.0/index.html>

NIH Public Access Policy Details <https://publicaccess.nih.gov/policy.htm>

NIH Public Access Policy. Submission Methods [https://publicaccess.nih.gov/submit\\_process.htm](https://publicaccess.nih.gov/submit_process.htm)

NISO JATS <http://jats.niso.org/>

NISO JATS Version 1.0 <http://jats.nlm.nih.gov/publishing/1.0/>

NLM Archiving and Interchange DTD, version 1.0. <http://dtd.nlm.nih.gov/archiving/1.0/>

NLM FTP Site <ftp://ftp.ncbi.nlm.nih.gov/pub/jats/>

NLM Journal Archiving and Interchange Tag Suite <http://dtd.nlm.nih.gov/>

NLM Journal Publishing DTD, version 1.0. <http://dtd.nlm.nih.gov/publishing/1.0/>

Modules in the JATS DTD Suite - Journal Article Tag Suite <http://jats.nlm.nih.gov/archiving/tag-library/1.1d1/n-mzy2.html>

Online PMC Style Checker <http://www.ncbi.nlm.nih.gov/pmc/tools/stylechecker/>

PLOS ONE <http://www.plosone.org/>

PLOS ONE Guidelines for Reviewers <http://www.plosone.org/static/reviewerGuidelines>

PLOS ONE Manuscript Guidelines <http://journals.plos.org/plosone/s/submission-guidelines>

pmc-1.dtd <http://www.ncbi.nlm.nih.gov/pmc/pmcdoc/dtd/index.cgi?show=../pmc-1.dtd>

SciELO model – Electronic Publication  
<http://scielo.org/php/level.php?lang=en&ycomponent=42&yitem=28>

SciELO PC Programs <http://docs.scielo.org/projects/scielo-pc-programs/en/latest/installation.html>  
[Update: Aug. 2015]

SciELO Project Documentation <http://docs.scielo.org/en/latest/>

SciELO Publishing Schema - Guia de uso dos elementos e atributos para documentos em XML  
[http://docs.scielo.org/projects/scielo-publishing-schema/pt\\_BR/latest/](http://docs.scielo.org/projects/scielo-publishing-schema/pt_BR/latest/)

SciELO Publishing Schema <http://docs.scielo.org/projects/scielo-publishing-schema/>

SicenceCentral <http://e-sciencecentral.org>

Standardized Markup for Journal Articles - National Information Standards Organization. Retrieved February 8, 2016, from <http://www.niso.org/workrooms/journalmarkup>

Style Checker <http://www.ncbi.nlm.nih.gov/pmc/pub/stylechecker-info/>

Tip: Flexible DTDs with parameter entities. Variable substitution in XML DTDs  
<http://www.ibm.com/developerworks/library/x-tipentdtd/>

XML Exchange Table Model Document Type Definition. Organization for the Advancement of  
Structured Information Standards (OASIS) <https://www.oasis-open.org/specs/tm9901.htm>

XML Information. The modules <http://www.ncbi.nlm.nih.gov/staff/beck/xml/index.html>





## ANEXO 1. FORMULARIO DE LA ENCUESTA

### Survey on XML usage - Digital Publishing Services provided by academic libraries

Answering these 5 questions will only take approximately 2-3 minutes of your time. Thank you in advance for your interest in this study. Your participation will contribute to a master's thesis research and also to the growth of knowledge in this area.

#### 1. Are you currently implementing XML in your library publishing activities?

Yes.

Not now, but we are planning to do it in the near future.

None of the above.

#### 2. If your answer to the first question is "none of the above", please specify the reason by choosing one or several of the following options to complete the sentence.

We are not interested in implementing XML in our publishing activities because...

our current publishing system doesn't include it as an option.

it is not required by our journal editors as an output.

we lack trained human resources on XML tagging.

it is too complicated to implement.

it is expensive to get started, use and maintain.

the added cost of XML markup doesn't pay off long term.

we see no advantages in using XML language.

Other reasons: \_\_\_\_\_

#### 3. In your opinion, which are the advantages of the adoption of XML format (application, vocabulary or language) in publishing services offered by libraries?

XML workflows enable library publishers to: (Please choose one or several of the following options).

output content quickly and easily in several electronic formats (e.g. ePub, HTML, PDF).

repurpose content into other channels (catalogs, websites, databases, printers).

automate processes.

scale publishing services and publications.

preserve the digital content for the future.

promote the visibility of the journals.

Other advantages: \_\_\_\_\_



The following questions apply only to those who actually implement or are planning to implement XML workflows.

**4. Which XML format (application, vocabulary or language) are you using/planning to use and why? (E.g. JATS, TEI, DB).**

**5. At what point do you integrate/are you planning to integrate XML into the workflow?**

Please choose one of the following options:

At the end of the publishing process, just before publishing the article.

At the middle of the publishing process.

All the way through the publishing process. Articles are submitted/uploaded to our system in XML.

Other: \_\_\_\_\_

Please feel free to include any other comments you would like to add about your publishing workflow.



## ANEXO 2. RESUMEN DE LAS RESPUESTAS

	1.1	1.2	1.3	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8*	3.1	3.2	3.3	3.4	3.5	3.6	3.7*	4.*	5.1	5.2	5.3	5.4*	Com*
EN01																								
EN02																								
EN03																								
EN04																								
EN05																								
EN06																								
EN07																								
EN08																								
EN09																								
EN10																								
EN11																								
EN12																								
EN13																								
EN14																								
EN15																								
EN16																								
EN17																								
EN18																								
EN19																								
EN20																								
EN21																								
EN22																								
EN23																								
EN24																								



	1.1	1.2	1.3	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8*	3.1	3.2	3.3	3.4	3.5	3.6	3.7*	4.*	5.1	5.2	5.3	5.4*	Com*
EN25																								
EN26																								
EN27																								
EN28																								
EN29																								
EN30																								
EN31																								
EN32																								
EN33																								
EN34																								
EN35																								
EN36																								
EN37																								
EN38																								
EN39																								
EN40																								
EN41																								
EN42																								
EN43																								
EN44																								
EN45																								

\*: La respuesta a esta opción o pregunta es abierta. Los textos correspondientes se presentan en las páginas siguientes.

## Textos correspondientes a las preguntas con opción a respuestas abiertas

### Pregunta 2, opción 8 "Other" (codificada como 2.8)

EN02	<i>Digital Commons uses OAI - XML and we rely on that output.</i>
EN05	<i>We would like to, it's just not as high on the priority list as some other things.</i>
EN19	<i>Not useful for multimedia publishing.</i>
EN38	<i>It hasn't come up yet.</i>
EN39	<i>We once helped a researcher get connected with a company that encodes journal articles according to PubMed standards.</i>
EN45	<i>We may implement xml in the future.</i>

### Pregunta 3, opción 7 "Other" (codificada como 3.7)

EN32	<i>Unsure.</i>
EN33	<i>Content harvesting.</i>
EN34	<i>Batch upload of backfiles.</i>
EN38	<i>Ability to submit to organizations that use OA such as PubMed.</i>
EN39	<i>I don't know enough about it to have an opinion - sorry!</i>
EN44	<i>Promote standards for data sharing.</i>

### Pregunta 4 (codificada como 4)

EN01	<i>Modified TEI is what we use. This is because our library publishing program originally focused on text encoding projects when it was started in 2002.</i>
EN08	<i>We currently use TEI for two of our journals, which are legacy projects. We hope to use JATS for any new journals that adopt XML publishing workflows.</i>
EN10	<i>I have no idea. XML formats are supplied by the vendor.</i>
EN14	<i>TEI.</i>
EN15	<i>We use the XML import feature to batch upload issues for adding back issues of a journal, or for journals who do peer review and editing external to OJS.</i>
EN20	<i>We haven't explored the options fully enough yet.</i>
EN21	<i>We currently use XML for metadata (not full text encoding) as part of the Open Journal Systems platform and in our institutional repository. In the future we may implement TEI for our monographs.</i>
EN22	<i>We use TEI for digital projects.</i>
EN28	<i>We convert ProQuest XML into Bepress Digital Commons XML.</i>
EN31	<i>JATS seems like the best option for journals, TEI for books. Mainly it will depend on what it supported best by the toolkits we will be using.</i>

EN33	<i>For export Dublin Core (DC), RIS for citation reference management and METS. Content is described in DC within repository.</i>
EN34	<i>In the past we have used TEI markup but have no plans to do this going forward with new collections/publications. We are using XML almost exclusively for batch uploading.</i>
EN36	<i>TEI for digital scholarship projects like Digital Thoreau (marking up text of Henry David Thoreau's Walden manuscripts).</i>
EN37	<i>We use journal article metadata in XML as part of our responsibility as CrossRef publisher members.</i>
EN40	<i>TEI, and may use others. The answers here apply to all of the digital production services operations, which include books, journals, and more. We have done less with our journals working with editors and publishers based on their interests, which are focused more on other areas including business models.</i>
EN41	<i>XML is generated automatically by Indesign.</i>
EN42	<i>We use JATS for the Stockholm University Press platform (powered by Ubiquity Press) see <a href="http://www.stockholmuniversitypress.se/site/research-integrity/#no-lockin">http://www.stockholmuniversitypress.se/site/research-integrity/#no-lockin</a></i>

**Pregunta 5, opción 4 "Other" (codificada como 5.4)**

EN01	<i>We have several different XML workflows for different product types.</i>
EN22	<i>We do not use XML for many aspects of publishing, i.e. ETDs, faculty journal articles in the IR etc. (all mainly PDF). We also use TEI for some ebook workflows.</i>
EN29	<i>We aren't currently planning to integrate.</i>
EN33	<i>Content is described in XML in the repository.</i>
EN34	<i>Post publication.</i>
EN38	<i>No plan at this time. We don't have the staff or the knowledge to be able to implement this. We are a small college.</i>
EN39	<i>If it becomes necessary/valuable.</i>
EN40	<i>Integrated with the repository system, and can be refined.</i>

**Pregunta 5, opción Comentarios (codificada como Com)**

EN40	<i>This relates to many production workflows. For more specifics, contact [nombre del contacto], Director of Digital Production Services, [correo-e del contacto].</i>
EN42	<i>Our systems support different file formats, but I do think that the XML files are created at the very end of the publishing process, as most files for peer-review are handled as word or pdf files. However, it might be that these files are converted to XML already at submission stage, but I'm not entirely sure about that. We are about to start using a newly built system for book submissions, so this routine may also change.</i>